

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:01:41 ; Search time 26.64 Seconds  
(without alignments)  
1915.784 Million cell updates/sec

Title: US-09-714-882-2  
Perfect score: 3570  
Sequence: 1 MPELSLLIEILILGVITKT.....LFGVLIPLVLLLRNHHG 689

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101:\*

1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*

22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3570	100.0	689	22	AAE02430 Novel human protei
2	3554.5	99.6	688	22	AAE02431 Novel human protei
3	2813	78.8	590	22	AAE02432 Novel human protei
4	2792.5	78.2	576	22	AAE02435 Novel human protei
5	2777	77.8	575	22	AAE02436 Novel human protei
6	2392	67.0	499	22	AAE02434 Novel human protei
7	2130	59.7	418	22	AAE02433 Novel human protei
8	1650	46.2	794	20	AA17750 Human pancreas-spe
9	1650	46.2	794	21	AAB41251 Human ORFX ORF1015
10	1646	46.1	766	20	AA18096 Full length mouse
11	1646	46.1	788	20	AA18095 Full length mouse

12	1150.5	32.2	574	20	AA18094 Partial Sel-1L pro
13	1098	30.8	404	20	AA18097 Human Sel-1L prote
14	435	12.2	640	21	AA195006 Human secreted pro
15	435	12.2	979	22	AA139116 Human polypeptide
16	435	12.2	1038	22	AA140902 Human polypeptide
17	373	10.4	833	18	AA127148 HMG-CoA reductase
18	300	8.4	301	21	AA166654 Membrane-bound pro
19	300	8.4	301	22	AA187534 Human PRO1063. Ho
20	300	8.4	301	22	AA165177 Human PRO4063 (UNQ
21	275	7.7	384	21	AA129891 Human secreted pro
22	275	7.7	384	21	AA129892 Human secreted pro
23	272	7.6	274	18	AA120582 H. pylori secreted
24	271.5	7.6	256	17	AA105197 Helicobacter pylori
25	268	7.5	54	22	AA120409 Peptide #6843 enco
26	268	7.5	54	22	AA134894 Peptide #8931 enco
27	267.5	7.5	256	19	AA110987 H. pylori ORF 06ee
28	267.5	7.5	256	19	AA189870 Antigen from clust
29	258.5	7.2	256	19	AA183350 H. pylori GHP0 646
30	258.5	7.2	256	20	AA189992 Expressed antigen
31	248	6.9	338	20	AA189928 Antigen 1 from clu
32	248	6.9	353	20	AA189824 Protein encoded by
33	248	6.9	355	19	AA110975 H. pylori ORF hp5p
34	238.5	6.7	185	18	AA120262 H. pylori secreted
35	238.5	6.7	185	18	AA124622 H. pylori secreted
36	226.5	6.3	290	18	AA155539 H. pylori ORF 02ge
37	219	6.1	285	21	AA120876 Arabidopsis thalia
38	219	6.1	285	21	AA153482 Arabidopsis thalia
39	219	6.1	351	21	AA120875 Arabidopsis thalia
40	219	6.1	351	21	AA153481 Arabidopsis thalia
41	219	6.1	356	21	AA120874 Arabidopsis thalia
42	219	6.1	356	21	AA153480 Arabidopsis thalia
43	218.5	6.1	291	20	AA189880 Antigen 1 from clu
44	218.5	6.1	291	20	AA189827 Protein encoded by
45	217.5	6.1	313	18	AA155443 H. pylori protein

ALIGNMENTS

RESULT	1
AAE02430	
ID	AAE02430 standard; Protein: 689 AA.
XX	
AC	AAE02430;
XX	
DT	10-AUG-2001 (first entry)
XX	
DE	Novel human protein (NHP) #1, sharing similarity with Notch ligand.
XX	
KW	Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW	novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW	polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW	pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW	Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW	cholesterol metabolism; coronary artery disease; gene therapy;
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Misc-difference 393
FT	/note= "Encoded by RAA"
XX	
PN	WO200136636-A2.
XX	
PD	25-MAY-2001.
XX	
PF	16-NOV-2000; 2000WO-US31373.
XX	
PR	17-NOV-1999; 99US-0165959.
XX	
PA	(LEXI-) LEXICON GENETICS INC.
XX	

PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX WPI; 2001-355635/37.  
DR N-PSDB; AAD06374.  
XX Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
PS Claim 2; Page 27-28; 39pp; English.  
XX  
CC The present sequence is a novel human protein (NHP) which share  
CC structural similarity with animal Notch ligands, particularly SEL-1.  
CC SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 689 AA;  
  
Query Match 100.08; Score 3570; DB 22; Length 689;  
Best Local Similarity 100.08; Pred. No. 2.1e-298;  
Matches 689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MKPLSLLEILLIILGVTIKTKAEHNKRQKERNVTTQVSNNEIKOYLSHILEQRTSSNV 60  
DB 1 mkplslleillilgvtiktkaehnrkrqernvttqvsneikqylshileqrtssnv 60  
  
QY 61 INKRENLEKKKNQKRIKIGTQNKDILKRNKNHLOKAEKNFTDEGDLFKMGIKVLIQQ 120  
DB 1 inkrenllekkknqrirkigtkqndilkrnknhlqkaeknftdegdlfkmgikvliqq 120  
  
QY 121 SKSQKQKEAYLLFAKADMGNLKAMEKADALLFCNFCVQNTAATOLYESLAKEGSCK 180  
DB 121 sxsqkqkeeyllfakadmgnlkamekadamallfngfvqntaailgylslakegsck 180  
  
QY 181 AQNALGFLSSYIGMEYDQAKALIYYTFCGAGNMMSQMLGYRSLGTVNLQNCVALS 240  
DB 181 aqnalgfllssyigmeydqakaliyytfcgagnmmsqmlgyrslgtnvlnqncvals 240  
  
QY 241 YKKKVADYIADTFEKSQGVPEKVKRLTERPENLSNSEILDWDIYQYKFLAERGDDVQIQ 300  
DB 241 ykkkvadyiadtfeksegvpvkekvrleterpenlsnselldwdiyykflaergdvgiq 300  
  
QY 301 VSLGOLHLIGRGLDQDYKALHVLKAAKAGSANAMAFICKMYLEGNAVPPQNNATAPK 360  
DB 301 vslgqlhligrkglldqdykalyhvlkkaagsanamafickmylegnavppqnnatafk 360  
  
QY 361 YFSMAASKGNATLHGLGLLYPHGKGVPLNYAEALKYFOKAEKGPDAQFQFGPMYYSG 420  
DB 361 yfsmaaskgnatlgglglyphgkgvplnyaealkyfokaekgpdaqfqlgfmymysg 420  
  
QY 421 SGIKWDYKLAIFYLASQSGOPLAIYYLAKMYATGTGTVRSCRTAVELYKGVCBLGHW 480  
DB 421 sgikwdyklafyiflasqsgoplaiyyalakmyatgtgtvrrscrtavelykgvcblghw 480  
  
QY 481 EKFLTAYAYKGDGDDSSIVQYALLAEMGYEVAQNSAFILESKANILEKMYPMALL 540  
DB 481 ekfltayaykdgddssivqyallaemgyevaqsnsafileskanilekmypmall 540

QY 541 LWNRAAIOGNAPARVIGDYHYGYGTTKDYOTAATHYSIAANKYHNAQAMENLAYMYEH 600  
DB 541 lwnraaiognafarvkgdyhygygtygtkdyotaathysiaaankyhaqamfnlaymyeh 600  
  
QY 601 GIGITKDIHLARRLYDMAAQTSDDAHIPVLFVAVMKLETHLLRDILFFNQFTRNNWLKL 660  
DB 601 gigitkdihlarrlydmaaqtsddahipvlfavmkletthllrdilffnqftrnnwkl 660  
  
QY 661 DNTIGPHWDLFVIGLIVPGLILLLNHHG 689  
DB 661 dntigphwldfvlglivpglllllnhng 689  
  
RESULT 2  
AAE02431  
ID AAE02431 standard; Protein: 688 AA.  
AC AAE02431;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Novel human protein (NHP) #2, sharing similarity with Notch ligand.  
XX  
KW Human; neurotropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 393  
FT /note= "Encoded by RAA"  
XX  
XX WO200136636-A2.  
XX 25-MAY-2001.  
XX 16-NOV-2000; 2000WO-US31373.  
XX 17-NOV-1999; 99US-0165959.  
XX (LEXI-) LEXICON GENETICS INC.  
XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX WPI; 2001-355635/37.  
XX N-PSDB; AAD06375.  
XX  
XX Novel isolated human polynucleotides encoding polypeptides sharing  
XX sequence similarity with mammalian SEL-1 proteins which are negative  
XX regulators of Notch family receptors, useful for treating diabetes,  
XX cancer  
XX  
XX Claim 3; Page 29-30; 39pp; English.  
XX  
XX The present sequence is a novel human protein (NHP) which share  
XX structural similarity with animal Notch ligands, particularly SEL-1.  
XX SEL-1 proteins are negative regulators of Notch family receptors.  
XX Notch receptors and their associated signalling pathways have been  
XX associated with development, apoptosis, neuron growth and maintenance.  
XX Labeled NHP probes can be used to screen a human genomic library which  
XX is helpful for identifying polymorphisms, determining the genomic  
XX structure of a given locus/allele and designing diagnostic tests. The  
XX NHP is also useful in screening techniques for drugs which treats  
XX symptomatic or phenotypic manifestations of perturbing the normal  
XX function of NHP in the body. Nucleotide constructs encoding functional  
XX NHPs, antisense, antisense molecules can be used in gene therapy  
XX approaches for modulating gene expression such as for preventing or

CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 688 AA;  
  
Query Match 99.68; Score 3554.5; DB 22; Length 688;  
Best Local Similarity 99.98; Pred. No. 4.6e-297;  
Matches 688; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
  
Qy 1 MKPLSLLEILLIGVTIKTIAEHNKRQERNVTTQVSVNEIKQYLSHILEQRTSSNV 60  
Db 1 mkplsllleilligvtiktiaeahnkrqernvttqsvneikqylshileqrtssnv 60  
  
Qy 61 INKRENLEKKKNQKIRIKIQNKDILKRNKHLQKAERNFTDEGDLFKMGIKVLIQQ 120  
Db 61 inkrenllekkknqrkirkigqnkdilkrnkhlgkaeknftdegdqlfmgikvliqq 120  
  
Qy 121 SKSOKOEEAYLLFPAKADMGNLKAMEKMDALLFGNFGVONITAAIOLYESLAKEGSK 180  
Db 121 sksokqeeayllfakaadmgnlkamekmdallfgnfgvqntaaiglyeslakegsk 180  
  
Qy 181 AQNALGFLSSYIGMEYDOAKALIYYTFGSAGGNMMSQMLGYRLSGINVLQCEVALS 240  
Db 181 aqnalgfllssyigmeydaqallytfgsaggnmsqmlgyrlysginvlqcevals 240  
  
Qy 241 YKKVADYIADTFEKGSEGVPEKVLTERPENLSNLSILDWDIYQYKFLAERGDVQIQ 300  
Db 241 ykkvadyiadtfekgsegvpekvrlterpenlsnlsnseildwdiyqykflaergdvqiq 300  
  
Qy 301 VSLGOLHLIGRGLDQDYKALHYFLKAAKAGSANAMAFICKMYLEGNAAVPONNATAFK 360  
Db 301 vslgqlhligrkgloddykalyflkalkaagsanamafickmylegnaavpnnatafk 360  
  
Qy 361 YFSAASKGNAIGLHGLGLYFHGKGVPLNYAEALKYFQKAAEKGPDPQALGFMYISG 420  
Db 361 yfsaaskgnaiglhlgllyfhgkgvplnyaealkyfqkaekgwpdaqqlgfmyisg 420  
  
Qy 421 SGIKWDYKLFKPYFLASQSQPLAIYLYLAKMYATGTVVRSCTAVELYKVCBLGHW 480  
Db 421 sgiwdyklfkyflasqsqplaiylakmyatgtvvrscrtavelykgvcbelghwa 480  
  
Qy 481 EKFLTAYFAYKDGDIIDSLVOYALLAEMGYEVAQNSAFILSKKANILEKEMYPMALL 540  
Db 481 ekfltayfaykdgdiidslvoyallaemgyevagnsafilskkanilekemypmall 540  
  
Qy 541 LWNRAAIOGNAPARVYKIGDYHYGYTKKDYQTAATHYSIAANKYHNAQAMFNLAYMEH 600  
Db 541 lwnraaiognafarvvykigydygytkkdyqtaathysiaankyhnaqamfnlaymeh 600  
  
Qy 601 GIGITKDIHLARLYDMAQSPDAHIPVFAVMKLETHLLRDILFFNQFTRWNWKL 660  
Db 601 gigitkdihlarlydmaqspdahipvfvamkletthllrdilffnqftrwnwkl 660  
  
Qy 661 DNTIGPHWDLFVIGLIVPGLILLRNHRG 689  
Db 660 dntigphwdlflviglivpglillrnhrhg 688  
  
RESULT 3  
AAE02432  
ID AAE02432 standard; Protein; 590 AA.  
XX  
AC AAE02432;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Novel human protein (NHP) #3, sharing similarity with Notch ligand.  
XX Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;

novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
Parkinson's disease; stroke; vascular dementia; fat metabolism;  
cholesterol metabolism; coronary artery disease; gene therapy;  
cerebroprotective.  
Homo sapiens.  
  
Key Location/Qualifiers  
Misc-difference 393 /note= "Encoded by RAA"  
  
WO200136636-A2.  
25-MAY-2001.  
  
16-NOV-2000; 2000MO-US31373.  
17-NOV-1999; 99US-0165959.  
(LEXI-) LEXICON GENETICS INC.  
Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
WPI; 2001-355635/37.  
N-PSDB; AAD06376.  
  
Novel isolated human polynucleotides encoding polypeptides sharing  
sequence similarity with mammalian SEL-1 proteins which are negative  
regulators of Notch family receptors, useful for treating diabetes,  
cancer  
  
Claim 4; Page 31-32; 39pp; English.  
  
The present sequence is a novel human protein (NHP) which share  
structural similarity with animal Notch ligands, particularly SEL-1.  
SEL-1 proteins are negative regulators of Notch family receptors.  
Notch receptors and their associated signalling pathways have been  
associated with development, apoptosis, neuron growth and maintenance.  
Labeled NHP probes can be used to screen a human genomic library which  
is helpful for identifying polymorphisms, determining the genomic  
structure of a given locus/allele and designing diagnostic tests. The  
NHP is also useful in screening techniques for drugs which treats  
symptomatic or phenotypic manifestations of perturbing the normal  
function of NHP in the body. Nucleotide constructs encoding functional  
NHPs, antisense, antisense molecules can be used in gene therapy or  
approaches for modulating gene expression such as for preventing or  
treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
insulinomas), blood pressure abnormalities, neurodegenerative diseases  
such as Parkinson's disease, stroke, vascular dementia and conditions  
requiring modulation of fat and cholesterol metabolism such as coronary  
artery disease.  
  
SQ Sequence 590 AA;  
  
Query Match 78.8%; Score 2813; DB 22; Length 590;  
Best Local Similarity 99.1%; Pred. No. 2.3e-233;  
Matches 549; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 MKPLSLLEILLIGVTIKTIAEHNKRQERNVTTQVSVNEIKQYLSHILEQRTSSNV 60  
Db 1 mkplsllleilligvtiktiaeahnkrqernvttqsvneikqylshileqrtssnv 60  
  
Qy 61 INKRENLEKKKNQKIRIKIQNKDILKRNKHLQKAERNFTDEGDLFKMGIKVLIQQ 120  
Db 61 inkrenllekkknqrkirkigqnkdilkrnkhlgkaeknftdegdqlfmgikvliqq 120  
  
Qy 121 SKSOKOEEAYLLFPAKADMGNLKAMEKMDALLFGNFGVONITAAIOLYESLAKEGSK 180  
Db 121 sksokqeeayllfakaadmgnlkamekmdallfgnfgvqntaaiglyeslakegsk 180

QY 181 AONALGFLSSYIGMEYDOAKALIYYTFTGSGAGNMMSOMILGYRYLSGINVLQNCCEVALS 240  
Db 181 aqnalglfssyigmeiydkaaliyytftgsagnmmsqmilgyrylsGINVLQNCCEVALS 240  
QY 241 YKKKVDYIADTFEKGSEGVPEKVRTERPENLSSNSETLDDWIIYQYKFLAERGDVQIQ 300  
Db 241 ykkkvadyiadtfekegsvpekvrtterpenlssnselldwiiyqykflaergdvqilq 300  
QY 301 VSLGQHLITGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAVPPQNNATAFK 360  
Db 301 vslgqlhlitgrkgldqdykalyhflkaakagsanamafigmylegnaavppqnnatafk 360  
QY 361 YFSMAASKGNAIGLHGLLPHGKGVPLNYAEALKYQKAEEKGWPDAQFOLGPMYVSG 420  
Db 361 yfsmaaskgnaiglhlgllyphgkgvplnyaealkyfqkaeekgwpdaqfqlgfmvysg 420  
QY 421 SGIKWDYKLAIFYLASOGOPALYIYLAQMYATGTGVVRSCTAVELYKVCGLGHW 480  
Db 421 sgiwkdyklaifylasogoplaalyilakmyatgtgvvrsctavelykgvcelghwa 480  
QY 481 EKFLTAYFAYKDGDDSSILVQYVALLAEMGYEVAQNSAFILSKKANILEKEKMPMALL 540  
Db 481 ekfltayfaykdgdssilvqyvalleaemgyevagqnsafilskkanilekekmypmall 540  
QY 541 LWNRAAIOGNAPAR 554  
Db 541 lwnraaigghslgq 554

## RESULT 4

AAE02435  
ID AAE02435 standard; Protein: 576 AA.

AC AAE02435;

DT 10-AUG-2001 (first entry)

DE Novel human protein (NHP) #6, sharing similarity with Notch ligand.

KW Human; neurotropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 393

FT /note= "Encoded by RAA"

XX WO200136636-A2.

XX 25-MAY-2001.

XX 16-NOV-2000; 2000WO-US31373.

XX 17-NOV-1999; 99US-0165959.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

DR WPI; 2001-355635/37.

DR N-PSDB; AAD06379.

XX Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer

XX

PS Disclosure: Page 36-37; 39pp; English.

XX The present sequence is a novel human protein (NHP) which share  
CC structural similarity with animal Notch ligands, particularly SEL-1.  
CC SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.

XX Sequence 576 AA;

Query Match 78.2%; Score 2792.5; DB 22; Length 576;

Best Local Similarity 81.0%; Pred. NO. 1.3e-231;

Matches 558; Conservative 7; Mismatches 11; Indels 113; Gaps 4;

QY 1 MKPLSLLEILLIGVTIKTAEHNKROKERNVTTQVSNVEIKQYLSHILEQRTSSNV 60

Db 1 mkplslleilligvtiktaehnkrokvntvtqvsneikqylshileqrtssnv 60

QY 61 INKRENLEKKKNQKIRIKGIONKDIILKRNKHQKAEKNFTDEGQDLFKMGIKVLQO 120

Db 61 inkrenllekkknqkirkigkionkdilkrnkhkqkaeknftdegdqlfmgikvllq 120

QY 121 SKSQKQEEAYLLFAKADMGNLKAMEKMAADALLFNGFVQNTAAIQLYESLAKEGSK 180

Db 121 sksqkqeeayllfakaadmgnlkamekmaadallfngfvqntaaiaqlyeslakegsk 180

QY 181 AQNALGFLSSYIGMEYDOAKALIYYTFTGSGAGNMMSOMILGYRYLSGINVLQNCCEVALS 240

Db 181 aqnalglfssyigmeiydkaaliyytftgsagnmmsqmilgyrylsGINVLQNCCEVALS 240

QY 241 YKKKVDYIADTFEKGSEGVPEKVRTERPENLSSNSETLDDWIIYQYKFLAERGDVQIQ 300

Db 241 ykkkvadyiadtfekegsvpekvrtterpenlssnselldwiiyqykflaergdvqilq 300

QY 301 VSLGQHLITGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAVPPQNNATAFK 360

Db 301 vslgqlhlitgrkgldqdykalyhflkaakagsanamafigmylegnaavppqnnatafk 360

QY 361 YFSMAASKGNAIGLHGLLPHGKGVPLNYAEALKYQKAEEKGWPDAQFOLGPMYVSG 420

Db 361 yfsmaaskgnaiglhlgllyphgkgvplnyaealkyfqkaeekgwpdaqfqlgfmvysg 420

QY 421 SGIKWDYKLAIFYLASOGOPALYIYLAQMYATGTGVVRSCTAVELYKVCGLGHW 480

Db 421 sgiwkdyklaifylasogoplaalyilakmyatgtgvvrsctave----- 468

QY 481 EKFLTAYFAYKDGDDSSILVQYVALLAEMGYEVAQNSAFILSKKANILEKEKMPMALL 540

Db 469 -----krtflkkr-----ci 480

QY 541 LWNRAAIOGNAPARVIGDYHYGYGCTKKDYQTAATHYSIAANKYHNAQAMENLAYMYEH 600

Db 481 qwr-----fsygielpf----- 492

QY 601 GLIGTKDIHLARRLYDMAAQTSPIAHIPVLFAVMKLETHLLRLDILFFNQFTTRWNWKL 660

Db 493 -----kdihlarryldmaaqtspiahipvlfavmklethllrldilffnqfttrwnwkl 547

QY 661 DNTIGPHWDLFVIGLIVPGLILLLNHHG 689



Db	548	dntigphwdlflvlgivpgllilrrnhhg	576
		Best Local Similarity 80.8%; Pred. No. 2.8e-230;	
		Matches 557; Conservative 7; Mismatches 11; Indels 114; Gaps	
RESULT	5		
AAE02436			
ID	AAE02436	standard; Protein; 575 AA.	
XX			
AC	AAE02436;		
DT	10-AUG-2001	(first entry)	
DE	Novel human protein (NHP) #7, sharing similarity with Notch ligand.		
XX	Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;		
KW	novel human protein; Notch ligand; apoptosis; neuron growth; therapy;		
KW	polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;		
KW	pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;		
KW	Parkinson's disease; stroke; vascular dementia; fat metabolism;		
KW	cholesterol metabolism; coronary artery disease; gene therapy;		
XX	cerebroprotective.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference 393		
FT	/note= "Encoded by RAA"		
PN	WO200136636-A2.		
XX			
PD	25-MAY-2001.		
XX			
PF	16-NOV-2000; 2000WO-US31373.		
XX			
PR	17-NOV-1999; 99US-0165959.		
XX			
PA	(LEXI-) LEXICON GENETICS INC.		
XX			
PI	Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;		
XX			
DR	WPI: 2001-355635/37.		
DR	N-PSDB; AAD06380.		
XX			
PT	Novel isolated human polynucleotides encoding polypeptides sharing		
PT	sequence similarity with mammalian SEL-1 proteins which are negative		
PT	regulators of Notch family receptors, useful for treating diabetes,		
PT	cancer -		
XX			
PS	Disclosure; Page 37-39; 39pp; English.		
XX			
CC	The present sequence is a novel human protein (NHP) which share		
CC	structural similarity with animal Notch ligands, particularly SEL-1.		
CC	SEL-1 proteins are negative regulators of Notch family receptors.		
CC	Notch receptors and their associated signalling pathways have been		
CC	associated with development, apoptosis, neuron growth and maintenance.		
CC	Labeled NHP probes can be used to screen a human genomic library which		
CC	is helpful for identifying polymorphisms, determining the genomic		
CC	structure of a given locus/allele and designing diagnostic tests. The		
CC	NHP is also useful in screening techniques for drugs which treats		
CC	symptomatic or phenotypic manifestations of perturbing the normal		
CC	function of NHP in the body. Nucleotide constructs encoding functional		
CC	NHPs, antisense, antisense molecules can be used in gene therapy		
CC	approaches for modulating gene expression such as for preventing or		
CC	treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,		
CC	insulinomas), blood pressure abnormalities, neurodegenerative diseases		
CC	such as Parkinson's disease, stroke, vascular dementia and conditions		
CC	requiring modulation of fat and cholesterol metabolism such as coronary		
CC	artery disease.		
XX			
SQ	Sequence 575 AA;		
Query Match	77.8%; Score 2777; DB 22; Length 575;		
Key	Location/Qualifiers		

FT Misc-difference 393 /note= "Encoded by RAA"  
 FT XX WO200136636-A2.  
 PN 25-MAY-2001.  
 XX 16-NOV-2000; 2000WO-US31373.  
 PD 17-NOV-1999; 99US-0165959.  
 PF (LEXI-) LEXICON GENETICS INC.  
 PR Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
 PA WPI: 2001-355635/37.  
 XX N-PSDB; AAD06378.  
 XX Novel isolated human polynucleotides encoding polypeptides sharing  
 PT sequence similarity with mammalian SEL-1 proteins which are negative  
 PT regulators of Notch family receptors, useful for treating diabetes,  
 PT cancer -  
 XX Claim 6; Page 34-35; 39pp; English.  
 PS The present sequence is a novel human protein (NHP) which share  
 CC structural similarity with animal Notch ligands, particularly SEL-1.  
 CC SEL-1 proteins are negative regulators of Notch family receptors.  
 CC Notch receptors and their associated signalling pathways have been  
 CC associated with development, apoptosis, neuron growth and maintenance.  
 CC Labeled NHP probes can be used to screen a human genomic library which  
 CC is helpful for identifying polymorphisms, determining the genomic  
 CC structure of a given locus/allele and designing diagnostic tests. The  
 CC NHP is also useful in screening techniques for drugs which treats  
 CC symptomatic or phenotypic manifestations of perturbing the normal  
 CC function of NHP in the body. Nucleotide constructs encoding functional  
 CC NHPs, antisense, antisense molecules can be used in gene therapy  
 CC approaches for modulating gene expression such as for preventing or  
 CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
 CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
 CC such as Parkinson's disease, stroke, vascular dementia and conditions  
 CC requiring modulation of fat and cholesterol metabolism such as coronary  
 CC artery disease.  
 XX Sequence 499 AA;  
 SQ

Query Match 67.0%; Score 2392; DB 22; Length 499;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-197;  
 Matches 468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPLSLLEILIIILGVTIKTAEEHNRKQKERNVTQVSNIEIKOYLSHILEQRTSSNV 60  
 DB 1 mkplslleiliiilgvtiktkaeehnrkqkernvtqvsneikqylshileqrtsnv 60

QY 61 INRENLEKKNQKIRIKIQNKDILKRNKHLQKQAEKNFTDEGDLFRMGIKVLQ 120  
 DB 61 inrenllekknqrkirkigiknkdilkrnkhlqkqae knftdegdlfrmgikvlq 120

QY 121 SKSOKQKEAYLLFAKADMGNLKAEKADALLFGVGNITAAIQLYESLAKESGCK 180  
 DB 121 sksqkqkeayllfakadmgnlkakekadal lfgvgvgnitaaiaqlyeslakegsc 180

QY 181 AQNALGFLSYGICMEYDQAKALIYYTFGSAGNNMNSQILGYRYLSGINVLQNCVVALS 240  
 DB 181 aqnalgfllsygicmeydqakaliyytfgsag nnnmsqilgyrylsginvlqncvals 240

QY 241 YKKVADYADTPEKSEGVPVEKVRTERPENLSSNSEILDWDIYYKFLAERGVDVQIQ 300  
 DB 241 ykkvadyadttekesegvpvekvrtterpen lssnselldwdiyykflaergdvqiq 300

QY 301 VSLGQLHLIGRKGLDDQYKALHYFLKAAKAGSANAMAFITGKMYLGSNAVPPONNATAFK 360  
 DB 301 vslgqlhligrkglddqykalhyflkaakag sanamafitgkmylgsnaavppnnatafk 360

Db 301 vslgqlhligrkglddqykalhyflkaakag sanamafitgkmylgsnaavppnnatafk 360  
 QY 361 YFSMAASKGNAIGLHGILGLLYFHGKGVPLNYAEALYFQKAAEKGWPDQAFOLGFMYYSG 420  
 Db 361 yfsmaaskgnaighlgilghlyfhgkgvplnyaeal yfkaaekgwpdqafqlgfmfysg 420  
 QY 421 SGIMKDYKLAFKYFYLASQSGQPLAIYYLAKMYATGTGVVRSRTAVE 468  
 Db 421 sgiwkdyklafkyfyiasqsgqplaiyy lakmyatgtgvvrsrtave 468

RESULT 7  
 AAE02433  
 ID AAE02433 standard; Protein; 418 AA.  
 XX  
 AC AAE02433;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Novel human protein (NHP) #4, sharing similarity with Notch ligand.  
 XX  
 KW Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
 KW cholesterol metabolism; coronary artery disease; gene therapy;  
 KW cerebroprotective.  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 393  
 FT /note= "Encoded by RAA"  
 XX WO200136636-A2.  
 XX 25-MAY-2001.  
 XX 16-NOV-2000; 2000WO-US31373.  
 XX 17-NOV-1999; 99US-0165959.  
 XX (LEXI-) LEXICON GENETICS INC.  
 XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
 XX WPI: 2001-355635/37.  
 XX N-PSDB; AAD06377.  
 XX Novel isolated human polynucleotides encoding polypeptides sharing  
 PT sequence similarity with mammalian SEL-1 proteins which are negative  
 PT regulators of Notch family receptors, useful for treating diabetes,  
 PT cancer -  
 XX Claim 5; Page 32-33; 39pp; English.  
 PS The present sequence is a novel human protein (NHP) which share  
 CC structural similarity with animal Notch ligands, particularly SEL-1.  
 CC SEL-1 proteins are negative regulators of Notch family receptors.  
 CC Notch receptors and their associated signalling pathways have been  
 CC associated with development, apoptosis, neuron growth and maintenance.  
 CC Labeled NHP probes can be used to screen a human genomic library which  
 CC is helpful for identifying polymorphisms, determining the genomic  
 CC structure of a given locus/allele and designing diagnostic tests. The  
 CC NHP is also useful in screening techniques for drugs which treats  
 CC symptomatic or phenotypic manifestations of perturbing the normal  
 CC function of NHP in the body. Nucleotide constructs encoding functional  
 CC NHPs, antisense, antisense molecules can be used in gene therapy  
 CC approaches for modulating gene expression such as for preventing or  
 CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
 CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
 CC such as Parkinson's disease, stroke, vascular dementia and conditions  
 CC requiring modulation of fat and cholesterol metabolism such as conditions  
 CC such as Parkinson's disease, stroke, vascular dementia and conditions

CC requiring modulation of fat and cholesterol metabolism such as coronary artery disease.

XX  
SQ Sequence 418 AA;

Query Match 59.7%; Score 2130; DB 22; Length 418;  
Best Local Similarity 100.0%; Pred. No. 8.3e-175;  
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKPLSLLEILLIGVTTIKTKAEHNNKROKERNVTVSVNNEIKOYLSHILEQRTSSNV 60  
DB 1 mkpLSllleIlllgvttIkTkaEhnnKrqkernvttqsvnneIkqYlshilleqRtssnv 60  
QY 61 INKRENLEKKKNRKIRIKGIONKDIILKRNKHLQKQAEKNFTDEGDLQFKMGIKVLQ 120  
DB 61 InKrenlleKkKnRkIrIkGIONkDiIlKrnKhlQkQaEknftdeGdlqfKmgIkVlq 120  
QY 121 SKSQKQKEEAYLLFKAADMGNLKAMEKMDALLFGNGVQVNTAAIQLYESLAKEGSK 180  
DB 121 sKsqKqKeEayllfKaAdmgNlKaMeKmdAlLfgNgvQvNtAAiQlYEsLaKeGsk 180  
QY 181 AQNALGFLSSYGIGMEYDQAKALIYYTFSAGGNMMSOMILGYRYLSGINVLQCEVALS 240  
DB 181 aQnaLgflSsyGigMeYdQaKaLiYyTfSagGnMmsOmIlGyRyLsGiNvlQceValS 240  
QY 241 YKKVADYIADTFERSEGVPEKVRTERPENLSSNSEILDWDIYQYKFLAERGQVQIQ 300  
DB 241 yKkvAdyIadTfERseGvPeKvRteRPeNLSSnSEilDwDiYqYkFlaErGqVqIQ 300  
QY 301 VSLQGLHIGRGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNATAFK 360  
DB 301 vSlqGlHlIgrGLdQdyKaLhYfLkAAkAGsANaMaFiGkMyLeGnaAvPqNnAtaFk 360  
QY 361 YFSMAASKGNATGLHGLGLLYFHGKGVPLNYAEALKYFQKAEKGPDAQFOLGFWY 418  
DB 361 yfSmaasKgnATglHglGLlyfHgKgvPlNyaeAlkYfQkAEkGpDaQfOLgfwY 418

RESULT 8  
AAV17750  
ID AAY17750 standard; Protein; 794 AA.  
XX  
AC AAY17750;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human pancreas-specific tumour suppressor TSA305 protein.  
XX  
KW Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;  
KW therapy; tumour suppressor.  
XX  
OS Homo sapiens.  
XX  
PN WO9928457-Al.  
XX  
PD 10-JUN-1999.  
XX  
PF 25-NOV-1998; 98WO-JP05306.  
XX  
PR 20-APR-1998; 98JP-0126803.  
PR 28-NOV-1997; 97JP-0343789.  
XX  
PA (SAKA ) OTSUKA PHARM CO LTD.  
XX  
PI Harada Y, Ozaki K;  
XX  
DR WPI; 1999-358128/30.  
DR N-PSDB; AAX76577, AAX76578.  
XX  
PT Tumour suppressor gene TSA305 expressed specifically in pancreas,  
PT useful for diagnosis and treatment of pancreatic cancer  
XX

PS Claim 1; Page 38-41; 54pp; Japanese.

XX The present sequence represents the protein from a gene designated  
CC TSA305, which is expressed specifically in pancreas tissue. The gene,  
CC sequences hybridizing with it, its expression product, and antibodies  
CC recognizing the expression product, are useful in the investigation,  
CC diagnosis, prevention and treatment of pancreatic cancer.

SQ Sequence 794 AA;

Query Match 46.2%; Score 1650; DB 20; Length 794;  
Best Local Similarity 52.3%; Pred. No. 4.5e-133;  
Matches 313; Conservative 118; Mismatches 158; Indels 10; Gaps 3;

QY 96 OKQAEKNFTDEGDLQFKMGIKVLQOSKQKQKEEAYLLFAKADMGNLKAMEKMDALLF 155  
DB 172 eeaakrrmqeaaemyqtgmKilngsnkkskreayrYlqkaasmhktalervsyalLf 231  
QY 156 GNFGVQNTAAIQLYESLAKEGSKAQNALGFLSSYGIGMEYDQAKALIYYTFSAGNM 215  
DB 232 gdyLpqnIqaaremfeklteegSpKgtalGflyasgIgvnssqakalYyvtfgalgnl 291  
QY 216 MSQMTILGYRYLSGINVLQCEVALSYKKVADYIADTFEKSEGVPEKVRTERPENLSS 275  
DB 292 IahmvIgyrywagIvlgscsalThyrIvanhvasdisltggsVvqrIrIpdvengpm 351  
QY 276 NSEILDWDIYQYKFLAERGQVQIQVLSGLHIGRGLDQDYKALHYFLKAAKAGSAN 335  
DB 352 nsgmleedliqYyKflaErgdVqGvIqLhLhgRgveqhqrAdYfnlaaagnsh 411  
QY 336 AMAFIGKMYLEGNAAVPQNNATAFKYFSMAASKGNAIGLHGLLYFHGKGVPLNYAEAL 395  
DB 412 amafigKmyseGsdIvPqsnetaLhyfKkaadmgnpvqsglGmaylygrgVqVnydlal 471  
QY 396 KYFQAAEKGPDAQFOLGFWYSGSGTWKDYKLAFFYLASQSGOPLAIYYLAKMYAT 455  
DB 472 KyfQaaEgGwVdgqlIgsmyngIgvkrDykqalkYfnlasqgghilafynlaqmhAs 531  
QY 456 GTGVVRSCTAVELYKGVCELGHWAEKFLTAIFYAYKDGIDISLSVQYALLAEMGYEVAQS 515  
DB 532 gtgvmsrchtavelfKncvcrgrwserImtaysyKdgdynaavIqyllaEggyevaqs 591  
QY 516 NSAFILESKKANILEKMYPMALLWNRAATOGNAFARVKTGYHYGYGKTKDYQTAA 575  
DB 592 naafildgreasIvgenetyPrallhwnraasgYtvarkIgdYhfYfgtdvYetaf 651  
QY 576 THYSTAANKYHNAQAMFNLAYWYEHGLGTDKDIHLARRLYDMAAQTSPDAHPVLFAYMK 635  
DB 652 thyrIaseqghsaqamfnIgmheXgIqkqdiHlakrafyDMAaeaspdagVpVfIalck 711  
QY 636 LETWTHLL---RDILFFNQFTTRWNWKLQNTIGPHWDLF---VIGLIVPGLILLRNHH 688  
DB 712 lgvvyfIqyiretnIrdmftq---ldmdqllgpewdYlmtIiallIgtvIayqrqh 766

RESULT 9  
AAB41251  
ID AAB41251 standard; Protein; 794 AA.  
XX  
AC AAB41251;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORF1015 polypeptide sequence SEQ ID NO:2030.  
XX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnery; antiporiatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

















Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	373	10.4	833	3	US-08-699-103B-8	Sequence 4, Appli
2	194	5.4	95	3	US-08-699-103B-8	Sequence 8, Appli
3	144	4.0	1751	4	US-09-136-574A-44	Sequence 44, Appli
4	127	3.6	592	2	US-08-599-171A-30	Sequence 30, Appli
5	127	3.6	592	2	US-08-646-590B-30	Sequence 30, Appli
6	127	3.6	592	3	US-09-069-226-30	Sequence 30, Appli
7	127	3.6	592	4	US-09-412-184-30	Sequence 30, Appli
8	118.5	3.3	1147	3	US-08-470-260-5	Sequence 5, Appli
9	118.5	3.3	1147	3	US-08-471-491-5	Sequence 5, Appli
10	118.5	3.3	1147	4	US-08-466-662-5	Sequence 5, Appli
11	118.5	3.3	3289	2	US-07-477-451-2	Sequence 2, Appli
12	113	3.2	689	1	US-07-766-351-5	Sequence 5, Appli
13	113	3.2	689	1	US-08-059-032-5	Sequence 5, Appli
14	113	3.2	689	5	PCF-US91-07290-5	Sequence 5, Appli
15	112	3.1	1388	2	US-08-685-576-4	Sequence 4, Appli
16	110	3.1	2285	4	US-09-308-375-2	Sequence 2, Appli
17	106	3.0	484	2	US-08-879-260-4	Sequence 4, Appli
18	106	3.0	484	3	US-09-231-529-4	Sequence 4, Appli
19	106	3.0	484	4	US-08-977-816-4	Sequence 4, Appli
20	106	3.0	981	2	US-08-649-046-2	Sequence 2, Appli
21	106	3.0	1388	2	US-08-685-576-1	Sequence 1, Appli
22	104.5	2.9	1354	3	US-08-685-871-2	Sequence 2, Appli
23	103	2.9	1079	3	US-09-058-489-22	Sequence 22, Appli
24	103	2.9	1240	3	US-09-058-489-23	Sequence 23, Appli
25	103	2.9	1347	3	US-09-058-489-24	Sequence 24, Appli
26	101.5	2.8	351	1	US-08-402-217A-2	Sequence 2, Appli
27	101.5	2.8	351	1	US-08-700-178-2	Sequence 2, Appli

Db 68 RNEKNYQSIWQNETDSQRHIYEL---LVQSEOFNNSEATYTLISQIHLWSQYNFPNHN 124  
QY 139 DMCNLKAMEKMDALLFGNFGVQNTAAIQLYESLAKESGCAQNALGPLSSYGVIMEYD 198  
Db 125 TIAH-KYLEKFNLDLTHFTWH-----SALFDLAVMATGCGASGNDQTVIPO-----D 170  
QY 199 QAKALIYTFGAGGNNMSQIMILGYRLSGINVLQNCVALSYKKVADYIADTFEKS- 257  
Db 171 SAKALIYQRAAQLGNLAKAKVLAYKYSYGFVPRNFHKSVLVYRDIAQLRKSYSROBW 230  
QY 258 -----GVPVEKVR-LTERPENLSNLSILWDIYQYK 289  
Db 231 DIVFPYTESYNRISDFESGLLKGNSVPSSTVRKRTTRPDIGS-----P 276  
QY 290 FLAERGDVOIQVS---LGOLHLGRKGL-----DQD-----YKALHVFELKA 331  
Db 277 FIAQVNGVOMTLOIEPMGRFAENGNDGNGINGEDEDEASERIRIYYAALNDY-KGIYS 335  
QY 332 GSAN-----AMAFIGKMYLEGNAAVPQNNATAF 359  
Db 336 QSRNCERAKNLLLELYKEFQPHVDNLDPLQVYVYVRCQLQLLGHMYFTGSGSKPNHMAE 395  
QY 360 KYP--SMAASKNAGLHG-----LGLLYFHGKGVPLNVAELKFKQA-----AEKGWD 408  
Db 396 EILTTSLEISR-RAQGPICRACIDLGLI---NQYITNNTISQAISYMKAMKTQANNG--I 449  
QY 409 AOPOLGFMYSGSGIWKDYKLA--FKYFVLASOSGPTAIYLYLAKMYATGTGVRSCTA 466  
Db 450 VEFQLSKLATS---FPEKIGDPNLMETAYLNGFIPAIYEFVAMIESGMNSKSVENT 505  
QY 467 VELYKGVCELGH--WAEKELTAYFAYKGDGIDSSIVOVALLAEMGYEVAQNSAFLESK 524  
Db 506 AYLKFTFVDKNEALMAPKLTAFALINDRSEVALWAYSQLAEQGVETAQVSAAYLMYQL 565  
QY 525 KANILE---KEKMPMALLNRAAIQGNAPARKYKIDYHYGYGTCKDYQTAATHYSI 580  
Db 566 PYEFEDPPRTDQRTKLTASYTRAFOKQGNIDAGVAGDIYF---OMQNYSKAMALYQG 621  
QY 581 AANKYHNAQAMFNLAYMYERHGLITKDIHLARRLYDMAAQTSPDAHIPVLFAVMKLETH 640  
Db 622 AALKY-SIQAINWNGYMHGHLGVNRDFHLAKRYDYQVSEHHRYLASKLSVLKL---H 677  
QY 641 LLRDILFFNQFTTRNWLKLDNTIGPHWD 669  
Db 678 LKSWLTWITR--EKVNYKWPSSPLPNED 704

RESULT 2  
US-08-699-103B-8  
; Sequence 8, Application US/08699103B  
; Patent No. 6107462  
; GENERAL INFORMATION:  
; APPLICANT: Rine, Jasper D.  
; APPLICANT: Hampton, Randolph  
; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/699,103B  
; FILING DATE: 16-AUG-1996  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/002,581  
; FILING DATE: 17-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Grant D.  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 09272/005001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/322-5070  
; TELEFAX: 650/854-0875  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-699-103B-8

Query Match 5.4%; Score 194; DB 3; Length 95;  
Best Local Similarity 60.0%; Pred. No. 2.4e-10;  
Matches 36; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 564 GYGTKKDTQTATHTSYIAANKYHNAQAMFNLAYMYERHGLITKDIHLARRLYDMAAQTSP 623  
Db 2 GLDQVDYETAPIHYRLASEQOHSQAAMFNLCYMHKEGLIGIKQDIIHLAKRFYDMAVSQP 61

RESULT 3  
US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1751 amino acids  
; TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 44:  
US-09-136-574A-44

Query Match 4.0%; Score 144; DB 4; Length 1751;  
Best Local Similarity 19.5%; Pred. No. 0.0009;  
Matches 106; Conservative 78; Mismatches 195; Indels 164; Gaps 25;

QY 140 MGNLAKAMEKADALLFGNFGVONITAAIOLYESLAKESCKAQNALGFLSSYGIGMEYDQ 199  
Db 1 MQEMKAIRK-----VVSIITALLVLTSLGCPGIMPVKAYAGTYNYGEALQ--- 46  
QY 200 AKALIYYFGSAGGNMMSOMILGYLGINVLQNCVALSY---YKVADYIADTFEKS 256  
Db 47 -RTIMEYEF-QMSGKLPKSVWRNWRGDSGLD--DGKDVCLDTGWHADGHHV----- 95  
QY 257 EGVPEKVLTERPENLNSNSILWDIYQYKFLAERGDVQIOLVSLGOLHLIGRGLDQ 316  
Db 96 -----KFNLP-----MSYSASMLGWAVEYKDAFVKSKOLE-----HILNQIEWAN 136  
QY 317 DYKKAH-----YFLKAKAGSANAMAFICKMYLE-----GNAAVPQNATAFK-----Y 361  
Db 137 DYFKCHPSKYIYYQ-----VGDPYVDHNFPGPAEVNMQMRPAYKCDLSNPA 184  
QY 362 FSWAASKGNAIGLHGLLPHGKGVPLNYAEALYFOKA-----AEKGWPDADFOLGF 415  
Db 185 SSVVETASLAVSVI-----KERNQAAAYLQHAQDLFFADTRSDAGYTAAT 237  
QY 416 MYTSGGIWKDYKLAFFKPYLASQSQPLAIYYLAKM-----YATGTGVVRS----- 463  
Db 238 GFYTSGETDGLGWAAMWLYIATNDS-----YLTAEELMSEYANGTWTQWDDVRY 292  
QY 464 -----RTAVELYKGYCE--LGHWAEK-----FLTAYFA 489  
Db 293 GTLIMAKITGELYKGAVERNLDHTDRTYTPKGMAYLTGWSLRVYATTAFLACVYA 352  
QY 490 KYDGDIDSLVYALLA--EMGYEVAQNSAFILSKKANILEKEMYPMALLNRAAI 547  
Db 353 DWSGCDSNKTKYLFNAKQIDYALGSTRSVVGF-----GTNYPQHPHHRNAHSS 404  
QY 548 QGNFARVKIGDYHY-----GYTKKDYQTAATHY-----SIAANKYHNAQAMFNLAY 596  
Db 405 WANS--MKIPEHRHILYGALVGGPGSDSYNDITDVOVEVACD--YNAGIVGALAK 459  
QY 597 MYE 599  
Db 460 MYQ 462

RESULT 4  
US-08-599-171A-30  
; Sequence 30, Application US/08599171A  
; Patent No. 5814473  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,171A  
; FILING DATE: Concurrently  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HERRON, CHARLES J.  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-38  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 592 AMINO ACIDS  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-599-171A-30

Query Match 3.6%; Score 127; DB 2; Length 592;  
Best Local Similarity 21.4%; Pred. No. 0.0055;  
Matches 120; Conservative 76; Mismatches 180; Indels 186; Gaps 28;  
QY 67 LLEKKKQKIRIKIGIONKDKILKR-----NKN--HLQKQAE----- 100  
Db 43 IVEKKKGKIRELVKALWGKDYKAKTGIGHTRWATHGKPTDENAHPTDEKGEFAVVHNGI 102  
QY 101 -KNFTDEGDQFKMGIKVLQSKSQKQKEAYLLFAK--AADM--GNLKAMEKMDALLF 155  
Db 103 IENYLEKEELKEGVK-----RSETDEVIAHLAKNVRGDLLEAVLTKVKKLGAF 158  
QY 156 GNFGVONITAAIOLYESLAKESCKAQNALGFLSSYGIGMEYDQAKALIYYTGSAGNM 215  
Db 159 -----AVITVHE-----PNRL-----IGVK--QGSPLI-VGLGEGENFL 189  
QY 216 MSQMLGYRLSGINVLQNCVALSYKKVADYIADTFE--KSEGVPEKVLTERPENL 273  
Db 190 ASDIPAILPYTKKIIVLDGE-----IADLTPTVNIYNFEGEPVSKYVN----- 234  
QY 274 SENSELWDIYQYKFLAERGDVQIOLVSLGOLHLIGRGLDQDYKALHYFLKA----- 328  
Db 235 -----ITPDLVS-----AEKG-----GEKHPMLKEIYEQP--KALNDILKGLSTE 274  
QY 329 -----AKGSANAMAFIKMWLEGNAAPQN--NATAFYKFSMAAS-KGN 370  
Db 275 DAIPFKLDFRRVLIITACGTSYHAGFVGKWTIERFAGVPTVIYASEFRYADVPSDKDI 334  
QY 371 ATGLHGLGLLYFHGKGVPLNYAEALYFOKAAEKGWPDADFOLGPMYSGSGTWKDYKLA 430  
Db 335 VIGISQSG-----ETADTKFALQSAKEG-----AFTVGLVNVVVGSAIDRE----- 375  
QY 431 FRYFYLASQSQPLAIYYLAKMWATGTGVVRSCTAVELYKGVCELGHWAEKFLTAIFY-- 488  
Db 376 -SDFSLSLTHAGP-----EIGVAATKTFTAQFTA 402  
QY 489 -----AYKGDIDSLVQYALLAEMGYEVAQS--NSAFILSKKANILEKEMYPMALLN 543  
Db 403 LYALSVRESEERENLIR--LLEKVPVSLVQTLNTABEEKVAEKYMKKNMNYLGRYLN 460  
QY 544 RAAIQGNAFARVKIGDYHYGY 565  
Db 461 PTALEG-ALKLKEISYIHAEGY 481

RESULT 5  
US-08-646-590B-30  
; Sequence 30, Application US/08646590B  
; Patent No. 5962283

GENERAL INFORMATION:  
APPLICANT: Warren, Patrick V.  
APPLICANT: Swanson, Ronald V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
CURRENT APPLICATION DATA:  
FASTSEQ for Windows Version 2.0  
APPLICATION NUMBER: US/08/646,590B  
FILING DATE: 08-May-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/599,171  
FILING DATE: 09-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/01094  
FILING DATE: 21-January-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/017001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 592 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-646-590B-30

Query Match 3.6%; Score 127; DB 2; Length 592;  
Best Local Similarity 21.4%; Pred. No. 0.0055;  
Matches 120; Conservative 76; Mismatches 180; Indels 186; Gaps 28;

QY 67 LLEKKKNQKIRIKGIONKDILKR-----NKN---HLQKQAE----- 100  
Db 43 IVEKKKGKIRELVKALMGKDYKAKTGHTWATHGKPTDENAHPTDEKGEFVAVHNGI 102

QY 101 -KNFTDEGOLFPMGKIVLQOSKOKOKEAYLLFAK--AADM--GNLKAMEKMDALLF 155  
Db 103 IENYLEKELKEKGVKF-----RSETDTEVIAHLIAKNGYDGLLEAVLTKVTKLKGAFAF 158

QY 156 GNFQGNITAAQLYESLAKESCKAQNALGFLSSVIGIGMEYDQAKALIYYTFGSAGGNM 215  
Db 159 -----AVITVHE-----PNRL-----IGVK-----QGSPLI-VLGEGENFL 189

QY 216 MSOMILGYRLSGINLQNCVALSYKKVADYIADTFE--KSEGVPEKVRILTERPENL 273  
Db 190 ASDIPAILPYTKIIVLDGE-----IADTPVTNVIYNEGEPVSKVM----- 234

QY 274 SSSNEILDNDIVQYKFLAERGVDQVLSGLHLIGRGLDQDYKALHYFLKA----- 328  
Db 235 -----ITPDLVS-----AEKG-----GPKHFLKKEIYEQP--KAINDTLKGLFSTE 274

QY 329 -----AKAGSANAMAFIGKMYLEGNAAPQN--NATAFYFNSAAS-KGN 370  
Db 275 DAIPFKLDFRVLIIACGTSYHAGEVGKWIYERFAGVPTVIYASEFRYADVPVSKDI 334

QY 371 AIGLHGLGLLYFHGKGVPLNYAEALKYQKAAEKGPDAQFQGLGPMYYSGSGIWKDYKLA 430  
Db 335 VIGISQSG-----ETADTKFALQSAKEG-----AFTVGLNVVVSADRE---- 375

QY 431 PKFYLASQSQGPLAIYYILAKMYATGTGVVRSCTRAVELYKGCVLGHWAEEKFLTAYP-- 488  
Db 376 -SDFSLTHAGP-----EIGVAATKTFTTAQFTA 402

QY 489 ---AYKGDGIDSSILVOYALLAEMGYEVAOS--NSAFILESKKANILEKEKMYPMALLWN 543  
Db 403 LYALSVRESEERENLR--LLEKVPSSLVEQTUNTAEBEVEKVAEKYMKKNMMLYLGRLNY 460

QY 544 RAAIQCNAAFARVKIGDYHYGY 565  
Db 461 PIALEG-ALKLKEISYIHAEGY 481

RESULT 6  
US-09-069-226-30  
; Sequence 30, Application US/09069226  
; Patent No. 6013509  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Patrick V.  
; TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/069,226  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/599,171  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HERRON, CHARLES J.  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-38  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 592 AMINO ACIDS  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; US-09-069-226-30

Query Match 3.6%; Score 127; DB 3; Length 592;  
Best Local Similarity 21.4%; Pred. No. 0.0055;  
Matches 120; Conservative 76; Mismatches 180; Indels 186; Gaps 28;

QY 67 LLEKKKNQKIRIKGIONKDILKR-----NKN---HLQKQAE----- 100  
Db 43 IVEKKKGKIRELVKALMGKDYKAKTGHTWATHGKPTDENAHPTDEKGEFVAVHNGI 102

QY 101 -KNFTDEGOLFPMGKIVLQOSKOKOKEAYLLFAK--AADM--GNLKAMEKMDALLF 155  
Db 103 IENYLEKELKEKGVKF-----RSETDTEVIAHLIAKNGYDGLLEAVLTKVTKLKGAFAF 158



```
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0316.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1147 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-260-5
```

Query Match 3.3%; Score 118.5; DB 3; Length 1147;

Best Local Similarity 19.6%; Pred. No. 0.094;

Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;

```
QY 16 VTIKTAEHNNKROKERNVTTQVSVNEIKOYLSHILEORTSSNVINKRENLEKKKNOR 75
   || || || || || || || || || || || || || || || || || || || || ||
Db 560 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 603

QY 76 KIRIKGIONKDILK--RNKNHLOKQAEKNFTDEGDQLFKMGIKVLOOSKSQKOEAYLL 133
   || || || || || || || || || || || || || || || || || || || || ||
Db 604 YDEVKQAK--KDLEKSLRKREHLEKEVEKKLESKSGKNKNMEAK-----AQANSOKDEIFAL 658

QY 134 FAKAADWG-----NLKAMEK--MADALIFGFGVQNTAAIQLYESLAKEGSKCAQNA 184
   || || || || || || || || || || || || || || || || || || || || ||
Db 659 INKEANRDARAIAAQNKGIRKELSDKLENVKNLKDQKDFDKSDFEFGKNKNDFSKAEET 718

QY 185 L-----GFLSSYGIGMEYDQAKALIYYTTFGSAGGNMMSOMILGYRLSGINVLNCEVALS 240
   || || || || || || || || || || || || || || || || || || || || ||
Db 719 LKALKGSVKDLGINPEW-----ISKVENLNAALN 747

QY 241 YK-----KVADYIADTFEKGSEGVPEKVELTERPENLSSNSELWDIYQYKFLA 292
   || || || || || || || || || || || || || || || || || || || || ||
Db 748 EFGNGKNKDFSKYTOAKSDLENSVKDVIINQ--KVTDKVDNLN-----QAVSVAK 795

QY 293 ERGDV--QIOVSLGQLHLGRKGLDQDYKALHYFLKAAKAGSANAMAFICKMY-----344
   || || || || || || || || || || || || || || || || || || || || ||
Db 796 ATGDFSRVEQALADLNFSKEQLAQ-----QAQKNESLNARK--KSEIYQSVKNGV 844

QY 345 ---LEGNAAPQNNATAFYKFS-----MAASKGNAIGLHGLGLLYFHGKVPPLNYAEALK 396
   || || || || || || || || || || || || || || || || || || || || ||
Db 845 NGTLVGNLSQAEATTLTKNFSDIKKELNAKLGNNNNNNNGL-----KNEPI-YAKVKN 898

QY 397 -----YFOAAEKGWPDQAFQFGMYYSGSGIKWDYKLAFFYFLASQSQGPL 444
   || || || || || || || || || || || || || || || || || || || || ||
Db 899 KKAGQAASLEPIYAQVAKK--VNAKIDRLNQIATASGLV-----VQQAAGFPPL 944

QY 445 ----ATYYLAKMYATGTGVVRSCTAVELYKGYCELGHWAEEKFLTAYFAYKDGID- 496
   || || || || || || || || || || || || || || || || || || || || ||
Db 945 KRHDKVDLSKV-----GLSRN---QELAQKIDNLNQAVSEAKAGFFGNLEQTDIKLKD 995

QY 497 -----SSLVQVALLAEMGYEVAQSNFATLESKKANILEK-----EKWYPMA 538
```

```
Db 996 STKHNPMLNWEASAKKVPASLSAKLDNYATNSHIRINSIKNGAINKATGMLTKNPEW 1055
QY 539 LLLNWRRAAIOGNAFARVKIGDYHYGYGPK--KDYQTAATHYIAANKYHNA-----588
   || || || || || || || || || || || || || || || || || || || || ||
Db 1056 LKLVNDKIVAHNV--GSVPLSEYDKIGFNOKNMDKSDSKF-----STKLLNAVKTNSGF 1110
QY 589 ----QAMFNLAYMY-----EHL 602
   || || || || || || || || || || || || || || || || || || || || ||
Db 1111 TQFLTNAFSTASYCLARENAEHGI 1135
```

#### RESULT 9

```
US-08-471-491-5
; Sequence 5, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-471-491-5
```

Query Match 3.3%; Score 118.5; DB 3; Length 1147;  
Best Local Similarity 19.6%; Pred. No. 0.094;  
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;

```
QY 16 VTIKTAEHNNKROKERNVTTQVSVNEIKOYLSHILEORTSSNVINKRENLEKKKNOR 75
   || || || || || || || || || || || || || || || || || || || || ||
Db 560 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 603

QY 76 KIRIKGIONKDILK--RNKNHLOKQAEKNFTDEGDQLFKMGIKVLOOSKSQKOEAYLL 133
   || || || || || || || || || || || || || || || || || || || || ||
Db 604 YDEVKQAK--KDLEKSLRKREHLEKEVEKKLESKSGKNKNMEAK-----AQANSOKDEIFAL 658

QY 134 FAKAADWG-----NLKAMEK--MADALIFGFGVQNTAAIQLYESLAKEGSKCAQNA 184
   || || || || || || || || || || || || || || || || || || || || ||
Db 659 INKEANRDARAIAAQNKGIRKELSDKLENVKNLKDQKDFDKSDFEFGKNKNDFSKAEET 718

QY 185 L-----GFLSSYGIGMEYDQAKALIYYTTFGSAGGNMMSOMILGYRLSGINVLNCEVALS 240
   || || || || || || || || || || || || || || || || || || || || ||
Db 719 LKALKGSVKDLGINPEW-----ISKVENLNAALN 747

QY 241 YK-----KVADYIADTFEKGSEGVPEKVELTERPENLSSNSELWDIYQYKFLA 292
   || || || || || || || || || || || || || || || || || || || || ||
Db 748 EFGNGKNKDFSKYTOAKSDLENSVKDVIINQ--KVTDKVDNLN-----QAVSVAK 795

QY 293 ERGDV--QIOVSLGQLHLGRKGLDQDYKALHYFLKAAKAGSANAMAFICKMY-----344
   || || || || || || || || || || || || || || || || || || || || ||
Db 796 ATGDFSRVEQALADLNFSKEQLAQ-----QAQKNESLNARK--KSEIYQSVKNGV 844

QY 345 ---LEGNAAPQNNATAFYKFS-----MAASKGNAIGLHGLGLLYFHGKVPPLNYAEALK 396
   || || || || || || || || || || || || || || || || || || || || ||
Db 845 NGTLVGNLSQAEATTLTKNFSDIKKELNAKLGNNNNNNNGL-----KNEPI-YAKVKN 898

QY 397 -----YFOAAEKGWPDQAFQFGMYYSGSGIKWDYKLAFFYFLASQSQGPL 444
   || || || || || || || || || || || || || || || || || || || || ||
Db 899 KKAGQAASLEPIYAQVAKK--VNAKIDRLNQIATASGLV-----VQQAAGFPPL 944

QY 445 ----ATYYLAKMYATGTGVVRSCTAVELYKGYCELGHWAEEKFLTAYFAYKDGID- 496
```



Db 945 KRHDVDDLSKV-----GLSRN-----QELAQKIDNLNOAVSEAKAGFFGNLEQOTIDKLKD 995  
Qy 497 -----SSLVQYALLAEMGYEVAOSNSAFILSKKANILEK-----EKWYPM 538  
Db 996 STKHNPMLNWLWESAKKVPASLKDNTVATNSHIRINSINRANGAINEKATGMTOKNPEW 1055  
Qy 539 LLLWNRAAIOGNAPARVKIGDYHYGYGK--KYQTAATHYSIAANKYHNA-----588  
Db 1056 LKLVNDKIVAHNV-GSVPLSEYDKIGFNQKNMKDYSDSFKF-----STKLNNAVKDTNSGF 1110  
Qy 589 ----QAMENLAYMY-----EHGL 602  
Db 1111 TOFLTNASTASYCLARENAEHGI 1135  
RESULT 10  
US-08-466-662-5  
; Sequence 5, Application US/08466662B  
; Patent No. 6130059  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; APPLICANT: Bugnoli, Massimo  
; APPLICANT: Telford, John  
; APPLICANT: Macchia, Giovanni  
; APPLICANT: Rappuoli, Rino  
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And  
; FILE REFERENCE: CHIR0057  
; CURRENT APPLICATION NUMBER: US/08/466.662B  
; CURRENT FILING DATE: 1995-06-06  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1147  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-08-466-662-5  
Query Match 3.3%; Score 118.5; DB 4; Length 1147;  
Best Local Similarity 19.6%; Pred. No. 0.094;  
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;  
Qy 16 VTIKTIAEHNKROKERNVTQVSNEIKOYLSHILEQRTSSNVINKRENLLERKKKQNR 75  
Db 560 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 603  
Qy 76 KIRIKGIONKDLK--RNKNHLQKQAEKNFTDEGDLQFKMGIKVLQOQSKOKKEEAYLL 133  
Db 604 YDEVKKAQ-KDLEKSLRREHLEKEVERKLESKSGNKNKMEAK-----AQANSOKDEIFAL 658  
Qy 134 FAKAADMG-----NLKAMEK-MADALLFGNFGVQNTAAIQLYESLAKEGSCKAONA 184  
Db 659 INKEANRARAAYAOQLKGIKRELSKLENVNNKLAOFKDSFDFKNGKNKDFSAKET 718  
Qy 185 L-----GFLSSYGIGMEYDQAKALIYYTFAGSGNMMSOMILGYRYLSGINVLQNCVALS 240  
Db 719 LKALKGSVKDLGINPEW-----ISKVENLNAALN 747  
Qy 241 YK-----KVADYIADTEKSEGVPEKVRVLTTERPENLSSNEILDWDIYQYKFLA 292  
Db 748 EFKNGKNKDFSKVTQAKSDLSNVKDVITNQ-KVTDKVDNLN-----QAVSVAK 795  
Qy 293 ERGDV-QIOVSLGQLHIGRLGDQDYKALHYFLKAAGKAGSANAWAFIGWY-----344  
Db 796 ATGDFSRVEQALADLNKFSKQLAQ-----QAOKNESLNARK-KSEIYQSVKNGV 844  
Qy 345 ----LEGNAVPONNATAFKYFS-----MAASKGNAIGHGLGLYFHGKGVPLNYAEALK 396  
Db 845 NGTLVGNLSQAEATLTKNFSDIKKELNAKLGNNNNNNGL-----KNEPI-YAKVYNK 898  
Qy 397 -----YFOKAAEKGWPDQAFQFGFMYSGSGIWKDYKLAIFYLASQSGQPL 444

Db 899 KGAQASLEETIYAQVAK--VNAKIDRLNQLASGLV-----VGQAAGFPL 944  
Qy 445 ---AIYVLAKMYATGTGVRSCRTAVELYKGVCELGHWAEKFLTAYFAYKDDID----496  
Db 945 KRHDVDDLSKV-----GLSRN-----QELAQKIDNLNOAVSEAKAGFFGNLEQOTIDKLKD 995  
Qy 497 -----SSLVQYALLAEMGYEVAOSNSAFILSKKANILEK-----EKWYPM 538  
Db 996 STKHNPMLNWLWESAKKVPASLKDNTVATNSHIRINSINRANGAINEKATGMTOKNPEW 1055  
Qy 539 LLLWNRAAIOGNAPARVKIGDYHYGYGK--KYQTAATHYSIAANKYHNA-----588  
Db 1056 LKLVNDKIVAHNV-GSVPLSEYDKIGFNQKNMKDYSDSFKF-----STKLNNAVKDTNSGF 1110  
Qy 589 ----QAMENLAYMY-----EHGL 602  
Db 1111 TOFLTNASTASYCLARENAEHGI 1135  
RESULT 11  
US-08-477-451-2  
; Sequence 2, Application US/08477451  
; Patent No. 5928865  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477.451  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0335.002  
; TELEPHONE: 510-601-2708  
; TELEFAX: 510-655-3542  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3289 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-477-451-2  
Query Match 3.3%; Score 118.5; DB 2; Length 3289;  
Best Local Similarity 19.6%; Pred. No. 0.52;  
Matches 134; Conservative 98; Mismatches 246; Indels 207; Gaps 32;  
Qy 16 VTIKTIAEHNKROKERNVTQVSNEIKOYLSHILEQRTSSNVINKRENLLERKKKQNR 75  
Db 2091 LTTKGLSPQEAANKL-----IKDFLSS--NKELVGKTLNFKAVADAKNTGN 2134  
Qy 76 KIRIKGIONKDLK--RNKNHLQKQAEKNFTDEGDLQFKMGIKVLQOQSKOKKEEAYLL 133  
Db 2135 YDEVKKAQ-KDLEKSLRREHLEKEVERKLESKSGNKNKMEAK-----AQANSOKDEIFAL 2189

QY 134 FAKADMG-----NLKAMEK-MADALLFGNFGVQNTAATAIQIYESIAKESGCKAQA 184  
Db 2190 INKEARNARATAYAQNLKIKRELSDKLENVKNLKDQDFDKNGKNGKDFSKAET 2249  
QY 185 L-----GFLSSYGIGMEYDOAKALIIYTFGSAGNMMSOMILGYRLSGINVLONCEVALS 240  
Db 2250 LKALGKSVKDLGINPEW-----ISKVENLNAALN 2278  
QY 241 YK-----KVADYIADTEKSEGVPEVKVRLTERPENLSSNSELIDWDIYQYKFLA 292  
Db 2279 EPKNGKNDKFSKVTQAKSDLENSKVDIINQ-KVTDKVDNLN-----QAVSVAK 2326  
QY 293 ERGDV-QIOVSLGQLHIGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMY-----344  
Db 2327 ATGDFSRVQALADLKNFSKEQLAQ-----QAQKNESINARK-KSEIYQSVKNGV 2375  
QY 345 ---LEGNAVPONNATAFKYS-----MAASKGNAIGLHGLGLYFHGKGVPLNTAEALK 396  
Db 2376 NGTLVGNGLSQAEATTLKSNFSDIKKELNAKLGNNNNNGL-----KNEPI-YAKVVK 2429  
QY 397 -----YFOKAAEKGPDAQFQGLGEMVYSGGIWKDYKLAIFYYLASOGPL 444  
Db 2430 KQAGQAASLEEIYIAQVAK--VNKIDRLNQIASGLV-----VQQAAGFPL 2475  
QY 445 ---AIYLLAKMYATGTGVVRSCTAVELYKGVCELGHWAERFLTAYFAYKGDID-----496  
Db 2476 KRHDVDDLKSV-----GLSRN---QELAQKIDNLQAVSEAKAGFGNLTQIDKLD 2526  
QY 497 -----SSLVQVALLAEMGYEVAQSNFASFILESKANILEK-----EKMYPWA 538  
Db 2527 STKHNPMLNWEVSARKVPASLSAKLDNYATNSHIRINSIKNGAINEKATGMLTOKNPEW 2586  
QY 539 LLLNRAAIOGNAFARVKIGDYHYGYGPK--KDYQTAATHYSIAANKYHNA-----588  
Db 2587 LKLVNDKIYAHNV-GSVPLSEVDKIGFNQKNMKDISDFK-----STKLNNAVKTNSGF 2641  
QY 589 ----QAMFNLAYMY-----EHLG 602  
Db 2642 TQFLTNAFSTASYCLARENAEHGI 2666

## RESULT 12

US-07-766-351-5

; Sequence 5, Application US/07766351

; Patent No. 5292652

; GENERAL INFORMATION:

; APPLICANT: Sinha, Sukanto

; APPLICANT: Seubert, Peter A.

; APPLICANT: Dovey, Harry F.

; APPLICANT: McConlogue, Lisa C.

; APPLICANT: Little, Sheila P.

; APPLICANT: Johnstone, Edward M.

; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Athena Neurosciences, Inc.

; STREET: 800F Gateway Blvd.

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/766.351

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Murphy, Lisabeth Feix

; REGISTRATION NUMBER: 31547

; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 689 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-07-766-351-5

Query Match 3.2%; Score 113; DB 1; Length 689;

Best Local Similarity 18.6%; Pred. No. 0.13;

Matches 95; Conservative 86; Mismatches 201; Indels 128; Gaps 19;

QY 32 ERNVTVQSVNIEIKOYLSHILBERTSSNVINKR-----ENLEKKKNQKIRIKGIQNK 85  
Db 72 EYTVTVQVNILDFPHVSPSKDIRTASTADKLLSEFDVEMSMREDVYQRIYVLQEKVOK 131  
QY 86 DIL-----KRNKNHLQKQAEKNFTDEGDQLFKMGIKVLOOSQKQKEA 130  
Db 132 DSLRPEAARYLERLIKLRNGHLPLRPTQENI-----KRIKKLSLLCIDFNKNLNEDT 186  
QY 131 YLLFKAADMGN-----LKAMEKMAADALLFGNFGVQNTAATAIQIYESLAKE-----176  
Db 187 TFLPFTLOELGGLPEDFLNSLEKMED-----GKLKVTLYKYPFPLLLKCHVPETRRK 239  
QY 177 -----GSKCAQNALCFLSSYGIGMEYDOAKALIIYTFGSAGNMMSOMILGYRLSGIN 230  
Db 240 VEEAENCRCKEENCA--ILKELVTLRAQKSRLLGLFTHAD-----YVLEMN 283  
QY 231 VLQNCVLSYKVKVADYIADTFEKGVPVEKVRLETERPENLSSNSELIDWDIYQY--288  
Db 284 MAKTSQTVATFLDELAQKLPLGEQERAVILELKRAECERRGLPFDGRITRADM--RYYN 342  
QY 289 -----KFLAERGDVOIQVSLGOL-----HLIGRKLGDQDYKALHYFLKAAKAGS 333  
Db 343 QVEETRYCVQNLKLEYFPVQV-VTHGLLGIYOELLGFLAFHHEEGASAWHEDVRLYTARD 401  
QY 334 ANAMAFIGKMYLE-----GNAA-----VFQNNATAFKYFSAAS-----KGNATG 373  
Db 402 AASGEVWGKFYLDLYPRECKYGHAAACFGLOPCLQDGSQRIATAIAAMVANFTKPTADAPS 461  
QY 374 L--HGLGLLYFHGKGVPLNYAEALKYFQKAAEKGPDAQFQGLGFMVYSGGIWKDYKFLAF 431  
Db 462 LLQHDVEVETFFH-----EFGHVMHQLCSQAE-----FAMFSGTHVERDFVEAP 504  
QY 432 KYFYLASOGSOPLAIFYLLAKMYATGTGVVR 461  
Db 505 SOM-LENNVWVQEPPLLRMSRHYRTGSAVPR 533

## RESULT 13

US-08-059-032-5

; Sequence 5, Application US/08059032

; Patent No. 5424205

; GENERAL INFORMATION:

; APPLICANT: Sinha, Sukanto

; APPLICANT: Seubert, Peter A.

; APPLICANT: Dovey, Harry F.

; APPLICANT: McConlogue, Lisa C.

; APPLICANT: Little, Sheila P.

; APPLICANT: Johnstone, Edward M.

; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend Kourie and Crew

; STREET: One Market Plaza, Steuart Street Tower,

; STREET: Suite 2000

CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 19930507  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, William E.  
REGISTRATION NUMBER: 30,223  
REFERENCE/DOCKET NUMBER: 15270-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 326-2400  
TELEFAX: (415) 326-2422  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 689 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-059-032-5

Query Match 3.2%; Score 113; DB 1; Length 689;  
Best Local Similarity 18.6%; Pred. No. 0.13;  
Matches 95; Conservative 86; Mismatches 201; Indels 128; Gaps 19;  
QY 32 ERNVTQSVNEIKQYLHLEORTSSNVINKR-----ENLEKKKNKIRIKIQNK 85  
Db 72 EYTYTVORNILDFPHQVSPSKDIRTASTADKKLSEFVEMSMREDVYQRIWVLEQVKQ 131  
QY 86 DIL-----KRNKNHLOKQAEKNFTDEGDLFKMGIKVLOQSKSQKQKEA 130  
Db 132 DLSRPEAARYLERLIKLGRRNGHLHLPRETQENI-----KRIKKLSLLCIDFNKLNEDT 186  
QY 131 YLLFAKAADMGN-----LKAMEKMDALLFGNFGVQNIITAAIOLYESLAKE----- 176  
Db 187 TELPFTLQELGGLPEDFLSLEKMD-----GKLKVTLYKYPHYPLKCKHVPETRRK 239  
QY 177 -----GSKCAQNALGFLSSYGIGMEYDQAKALIYTFGSAGNMMSQMLGYRYLSGIN 230  
Db 240 VEEAFNCRKEENCA--ILKELVTLRAQKSRLLGFTHAD-----YYLEMN 283  
QY 231 VLQNCVALSYKKVADYIADTFEKSEGVPEKVRITERPENLSSNSELIDWDIYQY-- 288  
Db 284 MAKTSOTVATFDELAKLPLCEQERAVILELKRAECERRGLPFDGRIRAWDM-RYYMN 342  
QY 289 -----RFLAERGDOVQVSLGQL-----HLIGRGLDQDYKALHYFLKAAKAGS 333  
Db 343 QVEETRYCVDQNLKKEYFPQV-VTHGLLGIYQELLGLAFHHEEGASAHWEDVRLYTARD 401  
QY 334 ANMAFIGKMYLE-----GNAA-----VPONNATAFKYFSAAS-----KGNATG 373  
Db 402 AASGEVVGKRYLDLYPREGKYGHAACGLQPGCLRGDSGRQIAAAMVANFTKPTADAPS 461  
QY 374 L--HGILGLLYFKGKGYPLNVAEALKYFQKAAEKGWDAQOFLGFMWYSSGIGWKDYKLA 431  
Db 462 LIQHDEVEYTFH-----BFGHVMHOLCSOAE-----FAMFSGTHVERDEVEAP 504  
QY 432 KYFYLASQSQGLAIYYLAKMYATGTGVVR 461  
Db 505 SQM-LENWWEQEPPLRMSRHYTGSAVPR 533

RESULT 14

PCT-US91-07290-5  
; Sequence 5, Application PC/TUS9107290  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/07290  
; FILING DATE: 19911004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Feix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 689 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PCT-US91-07290-5

Query Match 3.2%; Score 113; DB 5; Length 689;  
Best Local Similarity 18.6%; Pred. No. 0.13;  
Matches 95; Conservative 86; Mismatches 201; Indels 128; Gaps 19;  
QY 32 ERNVTQSVNEIKQYLHLEORTSSNVINKR-----ENLEKKKNKIRIKIQNK 85  
Db 72 EYTYTVORNILDFPHQVSPSKDIRTASTADKKLSEFVEMSMREDVYQRIWVLEQVKQ 131  
QY 86 DIL-----KRNKNHLOKQAEKNFTDEGDLFKMGIKVLOQSKSQKQKEA 130  
Db 132 DLSRPEAARYLERLIKLGRRNGHLHLPRETQENI-----KRIKKLSLLCIDFNKLNEDT 186  
QY 131 YLLFAKAADMGN-----LKAMEKMDALLFGNFGVQNIITAAIOLYESLAKE----- 176  
Db 187 TELPFTLQELGGLPEDFLSLEKMD-----GKLKVTLYKYPHYPLKCKHVPETRRK 239  
QY 177 -----GSKCAQNALGFLSSYGIGMEYDQAKALIYTFGSAGNMMSQMLGYRYLSGIN 230  
Db 240 VEEAFNCRKEENCA--ILKELVTLRAQKSRLLGFTHAD-----YYLEMN 283  
QY 231 VLQNCVALSYKKVADYIADTFEKSEGVPEKVRITERPENLSSNSELIDWDIYQY-- 288  
Db 284 MAKTSOTVATFDELAKLPLCEQERAVILELKRAECERRGLPFDGRIRAWDM-RYYMN 342  
QY 289 -----RFLAERGDOVQVSLGQL-----HLIGRGLDQDYKALHYFLKAAKAGS 333  
Db 343 QVEETRYCVDQNLKKEYFPQV-VTHGLLGIYQELLGLAFHHEEGASAHWEDVRLYTARD 401







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:05:26 ; Search time 13.36 Seconds  
(without alignments)  
1890.873 Million cell updates/sec

Title: US-09-714-882-2

Perfect score: 3570

Sequence: 1 MKPLSLLEILILGVTKT.....LFVIGLIVGLILLRLNRHG 689

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1650	46.2	794	1	SEIL_HUMAN
2	1635	45.8	790	1	SEIL_MOUSE
3	314	8.8	327	1	YBEQ_ECOLI
4	219	6.1	696	1	SRT5_YEAST
5	180	5.0	932	1	YAI1_SCHPO
6	179	5.0	515	1	Y141_HUMAN
7	168.5	4.7	250	1	HCPA_HELPY
8	161.5	4.5	250	1	HCPA_HELPJ
9	158.5	4.4	184	1	YBET_ECOLI
10	141.5	4.0	567	1	YES6_YEAST
11	137	3.8	512	1	YES6_YEAST
12	136	3.8	2869	1	REPL_PLAVB
13	133	3.7	267	1	EXOR_RHILV
14	132.5	3.7	165	1	YG25_HAETN
15	130.5	3.7	334	1	Y798_METJA
16	126.5	3.5	1150	1	YKDI_CABEL
17	125.5	3.5	268	1	EXOR_RHIME
18	125	3.5	1694	1	CLH1_DICDI
19	124	3.5	591	1	GLMS_AQUAE
20	122.5	3.4	3911	1	AKA9_HUMAN
21	119.5	3.3	723	1	DPD3_DROME
22	118.5	3.3	724	1	HMMR_HUMAN
23	118.5	3.3	1147	1	CGAI_HELPY
24	116	3.2	716	1	YOGF_BAGSU
25	114	3.2	864	1	YGM3_YEAST
26	114	3.2	945	1	ANPE_MOUSE
27	114	3.2	1235	1	DPOL_PYRHO
28	114	3.2	1738	1	YCF1_EPIVI
29	113	3.2	688	1	MEPD_HUMAN
30	112.5	3.2	514	1	TRPE_BUCRM
31	112.5	3.2	1286	1	RPOL_VARY
32	112.5	3.2	1699	1	DPOL_THEG8
33	112	3.1	1928	1	MYSL_YEAST

RESULT 1

ID	SEIL_HUMAN	STANDARD;	PRT;	794 AA.
AC	Q9UBV2; Q9P1T9; Q9UHK7;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	SEL-1 HOMOLOG PRECURSOR (SUPPRESSOR OF LIN-12-LIKE PROTEIN) (SEL-1L).			
GN	SEL1L OR TSA305.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreas;			
RX	MEDLINE=99426085; PubMed=10496078;			
RA	Harada Y., Ozaki K., Suzuki M., Fujiwara T., Takahashi E.-I.,			
RA	Nakamura Y., Tanigami A.;			
RT	"Complete cDNA sequence and genomic organization of a human			
RT	pancreas-specific gene homologous to Caenorhabditis elegans sel-1.";			
RL	J. Hum. Genet. 44:330-336(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Pancreas;			
RX	MEDLINE=20208354; PubMed=10746565;			
RA	Biunno I., Bernard L., Dear P., Cattaneo M., Volorio S., Zannini L.,			
RA	Bankier A., Zollo M.;			
RT	"SEL1L, the human homolog of C. elegans sel-1: refined physical			
RT	mapping, gene structure and identification of polymorphic markers.";			
RL	Hum. Genet. 106:227-235(2000).			
CC	-!- FUNCTION: MAY PLAY A ROLE IN NOTCH SIGNALING (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS.			
CC	-!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.			
CC	This SWISS-PROT entry is copyright.. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AB024763; BAA89204.1;			
DR	EMBL; AB024747; BAA89204.1; JOINED.			
DR	EMBL; AB024748; BAA89204.1; JOINED.			
DR	EMBL; AB024749; BAA89204.1; JOINED.			
DR	EMBL; AB024750; BAA89204.1; JOINED.			
DR	EMBL; AB024751; BAA89204.1; JOINED.			
DR	EMBL; AB024752; BAA89204.1; JOINED.			
DR	EMBL; AB024753; BAA89204.1; JOINED.			
DR	EMBL; AB024754; BAA89204.1; JOINED.			
DR	EMBL; AB024755; BAA89204.1; JOINED.			
DR	EMBL; AB024756; BAA89204.1; JOINED.			
DR	EMBL; AB024757; BAA89204.1; JOINED.			

```

DR EMBL; AB024758; BAA89204.1; JOINED.
DR EMBL; AB024759; BAA89204.1; JOINED.
DR EMBL; AB024760; BAA89204.1; JOINED.
DR EMBL; AB024761; BAA89204.1; JOINED.
DR EMBL; AB024762; BAA89204.1; JOINED.
DR EMBL; AB020335; BAA87904.1; -.
DR EMBL; AF052059; AAF29413.1; -.
DR EMBL; AF157516; AAF24176.1; -.
DR HSP; P02751; 2FN2.
DR MIM; 602329; -.
DR InterPro; IPR000562; FN_Type_II.
DR Pfam; PF00400; fn2; 1.
DR PRINTS; P00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00059; FN2; 1.
DR PROSITE; PS00023; FIBRONECTIN-2; 1.
KW Signal; Transmembrane; Glycoprotein.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 794 SEL-1 HOMOLOG.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 701 721 POTENTIAL.
FT TRANSMEM 739 759 POTENTIAL.
FT DOMAIN 127 168 FIBRONECTIN TYPE-II.
FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 186 186 M -> V (IN REF. 2).
FT SEQUENCE 794 AA; 88754 MW; 323EB03DC7485459 CRC64;

Query Match 46.2%; Score 1650; DB 1; Length 794;
Best Local Similarity 52.3%; Pred. No. 4.6e-100;
Matches 313; Conservative 118; Mismatches 158; Indels 10; Gaps 3;

QY 96 QXQAEKNFTDEGDLQFKMGKIVLQSKSQKQEEAYLLFAKADMGNLKAMEKMDALLF 155
DB 172 EBAARRRMOEAEEMMYQTKMTLINGSNKSQKREAYRYLQKAASNMHTKALERSVYALIF 231
QY 156 GNFGVQNTAAITOLVESLAKESCAQNAOAGFLSSVIGMEYDQAKALIYTFGSAGNM 215
DB 232 GYLQNTQAAEMEFKLEESGPKQTALGLFASGLGVNSQAKALYITFGALGML 291
QY 216 MSQMLGYRLSGINQVCEVALSYKVKVADYIADTEKSEGVPEVKVRLTERPENLSS 275
DB 292 IAHMVLGYRWAGIGVLOSCEALTHYRLVANHVASDISLTGSGVYQRLRPLDEVENPCM 351
QY 276 NSEILDNDIYQYKFLAERGDVQIQVSLGQLHLGRKLDQDYKALHYFLKAAKAGSAN 335
DB 352 NSGMLDEDLIQYQFLAERGDYQVQVGLGQLHLHGRGVQEQHRAFDFYFNLAANAGNSH 411
QY 336 AMAFTGMYLGNAAVPPONNATAFYESMAASKGAIGLGLGLYFFCKGVPYLAFAAL 395
DB 412 ANAFGLKMTSEGSIDVPSQNETALHYFKKAADMGPVGGSLGMAYLYGRGVQVNYDUAL 471
QY 396 KYFOAAERGWDAQFQLGFMYSIGTWKDYKLAFFYFLASQSQCPALYIYLAQMVA 455
DB 472 KYFOAAERGWDAQFQLGFMYSIGTWKDYKLAFFYFLASQSQCPALYIYLAQMVA 531
QY 456 GTGVVRSCTAVELKGVCELGHWAEKFLTAYFAYKDGDISLSVQYALLAEMGVEVAQS 515
DB 532 GTGVVRSCTAVELKGVCELGHWAEKFLTAYFAYKDGDISLSVQYALLAEMGVEVAQS 591
QY 516 NSAFILSKANILKEKMYPNALLWNRATQGNARVKIGDYHYGYGPKDYQATA 575
DB 592 NAAFILDQREASIGENETYPALHLWNRASQGTVTARKILGDFHFGFGFDVDETAF 651
QY 576 THYSTAANKYHNAQAMFNLAYMYEHLGITKDIHLARLYDMAAQTSPDAHIPVLFAVMK 635
DB 652 IHYRLASEQOHSQAAMFNGLYMHKGLGKIQDIHLAKRFYDMAEASPDQVPEVLALCK 711
QY 636 LETTHLL---RDILFFNQFTTRNNWKLKDNLTIGPHWDLF---VIGLIVPGLILLRNHH 688

Db 712 LCVWYFQVIRETNRDMFTQ-----LDMQLLPGPDWLYLMTIALLLTGTVIAYRQRQH 766

RESULT 2
SEIL_MOUSE
ID SEIL_MOUSE STANDARD; PRT; 790 AA.
AC Q942G6; Q9DBD8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SEL-1 HOMOLOG PRECURSOR (SUPPRESSOR OF LIN-12-LIKE PROTEIN) (SEL-1L).
GN SEIL OR SEL1H.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=99077704; PubMed=9858735;
RA Donoviel D.B., Donoviel M.S., Fan E., Hadjantonakis A.-K.,
RA Bernstein A.;
RT "Cloning and characterization of Sel-1l, a murine homolog of the C.
RT elegans sel-1 gene.";
RL Mech. Dev. 78:203-207(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Offelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: MAY PLAY A ROLE IN NOTCH SIGNALING (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF063095; AAD05210.1; -.
DR EMBL; AK005023; BAB23750.1; -.
DR MGD; MGI:1328016; Sellh.
DR ProDom; PD000995; FN_Type_II; 1.
KW Signal; Transmembrane; Glycoprotein; Alternative splicing.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 790 SEL-1 HOMOLOG.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 597 717 POTENTIAL.
FT TRANSMEM 735 755 POTENTIAL.

```









```
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: CONTAINS 5 TPR REPEATS.
CC -1- SIMILARITY: BELONGS TO THE HCPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF008565; AAB63298.1; -
DR EMBL; AF053708; AAC24210.1; -
DR EMBL; AE000541; AAD07279.1; -
DR TIGR; HP0211; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Signal; Repeat; TPR repeat; Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 250 HCPA PROTEIN.
FT REPEAT 29 62 TPR 1.
FT REPEAT 67 98 TPR 2.
FT REPEAT 100 133 TPR 3.
FT REPEAT 133 169 TPR 4.
FT REPEAT 169 202 TPR 5.
FT VARIANT 4 4 N -> S (IN ATCC 49503).
FT VARIANT 9 9 L -> F (IN ATCC 49503).
FT VARIANT 11 11 G -> W (IN ATCC 49503).
FT VARIANT 17 17 T -> A (IN ATCC 49503).
FT VARIANT 91 91 G -> S (IN ATCC 49503).
FT VARIANT 194 194 K -> Q (IN ATCC 49503).
FT VARIANT 199 199 K -> N (IN ATCC 49503).
FT VARIANT 235 235 S -> G (IN ATCC 49503).
SQ SEQUENCE 250 AA; 27365 MW; A375FF12EB795355 CRC64;

Query Match 4.7%; Score 168.5; DB 1; Length 250;
Best Local Similarity 23.3%; Pred. No. 0.00032;
Matches 60; Conservative 34; Mismatches 102; Indels 61; Gaps 6;
QY 340 IGKMYLEGNAAPV-----QNNATAFYFESMAASKGNAIGHLGLYFHG 384
Db 15 LGTILRLGLMAEPDAKELVNLGIESAKKQDFQAQKAFKACELKKEGF 74
QY 385 KGVLPLNVAEALYFQAAAEKGPDAQFOLGFMFYSGGIWKYKLAFFYLASQSQPL 444
Db 75 KGVGKDLKKAIOFYTKGCELNDGCGNLLGNLYNGOGYSKDAKASQYYSKACDLNHA 134
QY 445 AIYLLAKMYATGTVVRSCTAVELYKGVCGLGHWAKEKFLTAYFAYKD--GDIDSSLVQY 502
Db 135 GCMVLGSLHHYGVGTGPKDLRKALDLYEKACDL-----KDSPGCINAGYI-- 178
QY 503 ALLAEYGEVAQS--NSAFILSKKANILEKMKYPMALLLNRAAIOGNFAFVKIGDYH 561
Db 179 -----YSTKFKFAIVRYSKACELKDKRGCY-----NLGVMQ 211
QY 562 YYGYGTKKDYQTAAHY 578
Db 212 YNAQGTAKDEKQAVENF 228

RESULT 8
HCPA_HELPJ STANDARD; PRT; 250 AA.
AC Q9ZNM1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HCPA PROTEIN PRECURSOR (CYSTEINE-RICH 28 KDA PROTEIN).
GN HCPA OR JHP0197.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
```

```
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99120557; PubMed-9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 TPR REPEATS.
CC -1- SIMILARITY: BELONGS TO THE HCPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001458; AAD05781.1; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
KW Signal; Repeat; TPR repeat; Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 250 HCPA PROTEIN.
FT REPEAT 29 62 TPR 1.
FT REPEAT 67 98 TPR 2.
FT REPEAT 100 133 TPR 3.
FT REPEAT 133 169 TPR 4.
FT REPEAT 169 202 TPR 5.
SQ SEQUENCE 250 AA; 27294 MW; 94EC37877E98C578 CRC64;

Query Match 4.5%; Score 161.5; DB 1; Length 250;
Best Local Similarity 23.0%; Pred. No. 0.00092;
Matches 62; Conservative 36; Mismatches 106; Indels 65; Gaps 7;
QY 332 GSNAMAF-----IGKMYLEGNAAPV-----QNNATAFYFESMAASKGNAI 372
Db 3 GSVKTKLFLGVLGALCLRLGMAEPDAKELVSLGIESVKKQDFQAQKAFKACELKEGF 62
QY 373 GLHGLGLLYPHGKGVPLNVAEALYFQAAAEKGPDAQFOLGFMFYSGGIWKYKLAFK 432
Db 63 GCYFLGAFYEEGKGVGKDLKKAIOFYTKGCELNDGCGRLLGNLYNGOGYSKDAKASQ 122
QY 433 YFYLASQSGOPLAIYLLAKMYATGTVVRSCTAVELYKGVCGLGHWAKEKFLTAYFAYKD 492
Db 123 YYSKSELNHAECTVLGSLHHYGVGTGPKDLRKALDLYEKACDL-----KD 168
QY 493 --GDIDSSLVQYALLAEYGEVAQS--NSAFILSKKANILEKMKYPMALLLNRAAIOG 549
Db 169 SPCGINAGYM-----YGVAKNFEKAIYRYSKACELKDKRGCY----- 205
QY 550 NAFARVKIGDYHYGYGTTKDYQTAAHY 578
Db 206 -----NLGVMQYNAQGTAKDEKQAVENF 228

RESULT 9
YBET_ECOLI
ID YBET_ECOLI STANDARD; PRT; 184 AA.
AC P77296;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE HCPA PROTEIN PRECURSOR (CYSTEINE-RICH 28 KDA PROTEIN).
GN HCPA OR JHP0197.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
```

```
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Osefnet P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC CC -!- SIMILARITY: TO E.COLI YBEQ. SOME, TO YEAST SKT5 AND S.POMBE
CC SPAC24811.10C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000169; AAC73748.1; -
DR EMBL; U82598; AAB40848.1; -
DR EMBL; D90704; BAA35294.1; -
DR EMBL; D90705; BAA35299.1; -
DR Ecogene; Egl3650; yber.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 184 AA; 20915 MW; 7DC01030D545A0F9 CRC64;

Query Match
Best Local Similarity 4.4%; Score 158.5; DB 1; Length 184;
Matches 43; Conservative 34; Mismatches 55; Indels 23; Gaps 5;

QY 344 YLEGNAVPPQNATAFKYFSMAASKGNALGLHGLGLYFHGK-GVPLNVAEALKYFOKAA 402
DB 40 YLSSPGSLPFTTAAT-----ELSAIHGHTSQYRLGEFFLHSGDGKPLDYTQRYWEQSA 95
QY 403 EKGWPDAPQGLGFMYSSGIIWKYLFAPYLASQSGQPLAIYYLAKMYATGTGVVRS 462
DB 96 EOENPRAQSKLGWIKGLGVKPKDPRKAILWYKEAAEQGYAHAQYTLGLIYRNGSGI--- 152
QY 463 CHTAVELKGVCELGHW-AEKF--LTAFAYKGDG 494
DB 153 -----NVNHYESQKWLKLTAKOHYKNAE 175

RESULT 10
YE28_METJA STANDARD; PRT; 567 AA.
ID YE28_METJA
```

```
Q58823;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1428.
GN MJ1428.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 3661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Kleravage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC CC -!- SIMILARITY: CONTAINS 13 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67584; AAB99438.1; -
DR TIGR; MJ1428; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 9.
DR SMART; SM00028; TPR; 7.
KW Hypothetical protein; Repeat; TPR repeat; Complete proteome.
FT REPEAT 14 47
FT REPEAT 48 81
FT REPEAT 83 115
FT REPEAT 115 148
FT REPEAT 150 183
FT REPEAT 199 234
FT REPEAT 236 268
FT REPEAT 303 335
FT REPEAT 344 379
FT REPEAT 379 412
FT REPEAT 414 446
FT REPEAT 505 538
SQ SEQUENCE 567 AA; 66239 MW; 7CB4AC7210B7927E CRC64;

Query Match
Best Local Similarity 4.0%; Score 141.5; DB 1; Length 567;
Matches 112; Conservative 83; Mismatches 195; Indels 157; Gaps 27;

QY 22 KAEHNKROKERNVTQVSNEIKQYLHLEQRTSSNVINKRENLEKKKQKIRKG 81
DB 32 KAVECYLKALEKKNTPIDWFNLAVALYHLEKYDSALEAIN--EALKISPSNIYFAYLKG 89
QY 82 I---QNKDILKKNKHLOKQAEKNFTDE-----GQOLFPMGKIVLQQSQKQKEAYLL 133
DB 90 LIHYRGEITILAYK-YLKAKSEKINEELFEILDGTSVKYG-----RYEELKY 137
QY 134 FAKAADMGNLKAMEKMAA-----LLFGNFGVQNTAAITOLYESLAKEG---SCK-----A 181
DB 138 YLKSYSKMANSKNLNALFRAGKIYLLFG-----DIDKADAFNEILOQNPSCHECKIVECM 192
```



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M88097; AAA29743.1; --  
DR HSSP; P36956; IAM9.  
DR Malaria; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.  
FT DOMAIN 18 2807 EXTRACELLULAR.  
FT TRANSMEM 2808 2826 POTENTIAL.  
FT DOMAIN 2827 2869 CYTOPLASMIC.  
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;  
  
Query Match 3.8%; Score 136; DB 1; Length 2869;  
Best Local Similarity 20.0%; Pred. No. 1.2;  
Matches 155; Conservative 116; Mismatches 276; Indels 228; Gaps 36;  
  
QY 17 TTKTKAEHNKROKERNVTQVSNEIKOYLSHILEQRTSSNV-----INKRENLEKK 71  
DB 1218 TLKSNAGKGENHDELELNKTKGQMRDIYEKLTAEELKEGTVNELKDANEKANKVEPE 1277  
QY 72 NKQRKI-----RI-----KGIONKDIKLRN--KNHLOKQAEKFTDREGDOLFKMGI-KVLQQ 120  
DB 1278 PERNLIGHVLEIRTEKDKAGKVVEMNSLTKIKELKLOEVSDDSONELVTSITKHLEN 1337  
QY 121 SKS-----OKQKEAYLLFKAADMGNLKAMEKMDALLFGNFGVONTAATQIYSLAK 175  
DB 1338 AKGYEDVIKRNEDSIQLREKAKSLTLDKMKLQVQV-----NNNLQSAIQ-----GN 1386  
QY 176 EGSCAQNALGFLSYGMEYDQAKALIYYTFGSAGNWSQMILGYRYLSGINVLQNC 235  
DB 1387 AGISKELNELKGVIELLSTNY--SILEYVKKNSESVRFSQLANG-EFTKAEAGEERNA 1443  
QY 236 EVALGVYKVV-----ADY--IADTFKSEGVPEKVLTERPENLSNLSILDWDIYO 286  
DB 1444 SARLAEAEKLEQIVKDLIDYSDIDDKVKKIEGKIEILKME-----SALTWESE 1495  
QY 287 YK-----FLAERGD-VQIQVSLGHLIGRGLDQDYY-KALHYF- 325  
DB 1496 KFKQMCSSHMENAKGKKKIEYKLNKGCGKANITDSOMEVGV-----NYVSKAEHAFH 1549  
QY 326 -----LKAAGKAGSANAMAFIGKM-----YLEGNAAPQNNATAFKYFSMAASKGNAI 372  
DB 1550 TVEAQVDKTKAFCEIVAYVTVMKDNLFNLSLMKEVKVKCEKKNDEAEKYSK----- 1601  
QY 373 GLRGLGLLYFHGK--GVPLNYAEALKYFQKAAERGWPAQF-----QL 413  
DB 1602 -----LKPYDGRIKARVSENERKISELKEKAKVEKSSQLNDVSTKSLLOIDNCROQL 1655  
QY 414 GPMYSGGIWKDYKAFKYFYLASOSGOLAIYYLAKWYATGTGVVRSCTRAVELYKGV 473  
DB 1656 DSVL--SNIGRVKQNALQYFSDAKS-----MKSVLP 1686  
QY 474 CELGHWAKEFTLAYFAYKDG-DIDSSLYQYALLAEMGYEVAQNSAFILESK-----KANI 528  
DB 1687 SELG--AEKSLDKVKAAKESYEKNLETVO-----NEMSRINVEEGSLTDIDKKTITDIEND 1740  
QY 529 LEKEMYPMALLNRAATOQNA-----FARVKIGDYHYVG 564  
DB 1741 LAMKQYEEGLL-----QKIKENADRKSNFELVSGSEINALLDPSTISIFIKLKEYDWTG 1796  
QY 565 -----YGTK-----KDYOPAAATHYSTAAN--KYHNAQAMFNLAYWYHGLGITKDI 608  
DB 1797 DLKNYGVKMKNEIHEGFTKSYNLIETHLSNATDYSVTFEKAQSLRELAKEEE----- 1848  
QY 609 HLARLYDMAQTSDDAHPVLFAYWKLKLETHLLRDLFFNFQFTTRWNWKLKLDNT 663  
DB 1849 HLRRR-----EETAEIFLNDIKKVESLKLKEMM--KKVSAEYEGMKRDKHT 1892

RESULT 13  
EXOR\_RHLV STANDARD; PRT; 267 AA.  
AC Q52822;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE EXOPOLYSACCHARIDE PRODUCTION NEGATIVE REGULATOR PRECURSOR.  
GN EXOR.  
OS Rhizobium leguminosarum (biovar viciae).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=387;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WSM710 / WR6-35;  
RX MEDLINE=97346048; PubMed=9202471;  
RA Reeve W.G., Dilworth M.J., Tiwari R.P., Glenn A.R.;  
RT "Regulation of exopolysaccharide production in Rhizobium leguminosarum  
biovar viciae WSM710 involves exoR";  
RL Microbiology 143:1951-1958(1997)  
CC -!- FUNCTION: NEGATIVELY MODULATES EXOPOLYSACCHARIDE (EPS)  
BIOSYNTHESIS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L39937; AAB63941.1; -- Signal.  
KW Exopolysaccharide synthesis; POTENTIAL.  
FT SIGNAL 1 23  
FT CHAIN 24 267 EXOPOLYSACCHARIDE PRODUCTION NEGATIVE  
REGULATOR.  
SQ SEQUENCE 267 AA; 28922 MW; E45E2E0538E1F573 CRC64;  
  
Query Match 3.7%; Score 133; DB 1; Length 267;  
Best Local Similarity 32.5%; Pred. No. 0.073;  
Matches 49; Conservative 20; Mismatches 62; Indels 20; Gaps 6;  
  
QY 311 RGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVQNNATAFKYFSMAASKGN 370  
DB 55 KNGKEEAEVAYRAAEKHTGSRWALA---NNYADGD-GVTDQDDFEAFKIYSEIAQQGV 110  
QY 371 AIGLHGLGLL-----YFHG-KGVF-----LNYAEALKYFQKAAERGWPAQFQLGF 415  
DB 111 EPGSEDTGFFVNALLSLANYKHGIAGSPVRIDLQAROL-YFOVASTGVPFAQFOLAQ 169  
QY 416 MYYSGGIWKDYKAFKYFYLASQSQGLAI 446  
DB 170 MMLAGEGNGNASPQQAQKWLNLQARKSGHPGAM 200  
  
RESULT 14  
YG25\_HAEIN STANDARD; PRT; 165 AA.  
AC P44277;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE EXOPOLYSACCHARIDE PRODUCTION NEGATIVE REGULATOR PRECURSOR.  
GN EXOR.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:03:36 ; Search time 19.39 Seconds  
(without alignments)  
2706.770 Million cell updates/sec

Title: US-09-714-882-2  
Perfect score: 3570  
Sequence: 1 MKPLSLIILILGVTKT.....LFVIGLIVPGLILLRNHHG 689

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1159	32.5	685	2 T22223	sel-1 protein - Ca
2	723.5	20.3	604	2 E96762	hypothetical prote
3	455	12.7	700	2 T38028	hypothetical prote
4	391	11.0	807	2 T18454	hypothetical prote
5	373	10.4	833	2 S48558	probable membrane
6	314	8.8	327	2 B47999	hypothetical prote
7	275.5	7.7	713	2 T40051	hypothetical prote
8	267.5	7.5	256	2 G71856	hypothetical prote
9	258.5	7.2	256	2 E64659	conserved hypothet
10	255.5	7.2	268	2 T46587	hypothetical prote
11	248	6.9	355	2 G71958	hypothetical prote
12	247	6.9	477	2 H81845	hypothetical prote
13	244	6.8	355	2 C64549	conserved hypothet
14	240	6.7	468	2 F82012	hypothetical prote
15	228	6.4	195	2 F85564	hypothetical prote
16	228	6.4	467	2 B81095	conserved hypothet
17	226.5	6.3	290	2 C71859	hypothetical prote
18	225.5	6.2	290	2 B46657	conserved hypothet
19	219	6.1	696	2 S39827	SKT5 protein - yea
20	214	6.0	306	1 H64539	conserved hypothet
21	214	6.0	352	2 G71914	hypothetical prote
22	212	5.9	198	2 F81200	conserved hypothet
23	210.5	5.9	305	2 F71968	hypothetical prote
24	200.5	5.6	623	2 S67762	hypothetical prote
25	195.5	5.5	364	2 A36730	hypothetical prote
26	187	5.2	328	2 F83599	hypothetical prote
27	183.5	5.1	251	2 F71945	hypothetical prote
28	183.5	5.1	512	2 T41340	hypothetical prote
29	180	5.0	932	2 S62555	protoplast regene

30	176.5	4.9	225	2 D64598	hypothetical prote
31	170.5	4.8	321	2 T47138	hypothetical prote
32	169	4.7	271	2 F83081	hypothetical prote
33	168.5	4.7	250	2 C64546	conserved hypothet
34	168	4.7	2295	2 B71621	probable membrane
35	162	4.5	256	2 E71807	hypothetical prote
36	161.5	4.5	250	2 B71961	hypothetical prote
37	158.5	4.4	184	2 E84799	yber protein - Esc
38	156.5	4.4	184	2 B85565	hypothetical prote
39	153	4.3	633	2 T39352	hypothetical prote
40	153	4.3	1115	2 B84476	probable TPR repea
41	147	4.1	456	2 T40386	hypothetical prote
42	145.5	4.1	289	2 A83011	hypothetical prote
43	141.5	4.0	385	2 A82388	conserved hypothet
44	141.5	4.0	567	2 C64478	hypothetical prote
45	140.5	3.9	138	2 H64561	invalid gene - Hel

ALIGNMENTS

RESULT 1

T22223  
sel-1 protein - Caenorhabditis elegans  
N:Alternate names: regulator of lin-12  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T22223; S68303; S68304  
R:White, S.  
submitted to the EMBL Data Library, August 1996  
A:Reference number: Z19532  
A:Accession: T22223  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-685 <WIL>  
A:Cross-references: EMBL:Z78063; PIDN:CAB01505.1; GSPDB:GN00023; CESP:F45D3.5  
A:Experimental source: clone F45D3  
R:Grant, B.; Greenwald, I.  
Genetics 143, 237-247, 1996  
A:Title: The Caenorhabditis elegans sel-1 gene, a negative regulator of lin-12 and gl  
A:Reference number: S68303; MUID:96304591  
A:Accession: S68303  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-685 <GRA>  
A:Cross-references: GB:U50828; NID:g1255198; PIDN:AAC47112.1; PID:g1255199  
A:Accession: S68304  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-685 <GR2>  
A:Cross-references: EMBL:U50829; NID:g1255200; PID:g1255201  
C:Genetics:  
A:Gene: sel-1  
A:Map position: 5  
A:Introns: 41/1; 90/2; 210/3; 411/1; 548/1

Query Match 32.5%; Score 1159; DB 2; Length 685;  
Best Local Similarity 38.7%; Pred. No. 1.2e-64;  
Matches 253; Conservative 137; Mismatches 214; Indels 50; Gaps 11;

QY	53	EORTSSNVINKRENLEKKNKIRIKQIONKDILK-RNKNHLOKQAKNFTDEGDQLF	111
Db	57	EQVTSQDENK-----SNREIPKIVISEYLAEEKVEQPTSPAEAEF	97
QY	112	KMGIRVLOQSK--SQKQKEAYLLFAKADMGNLKAMEKMDALLFGNFGVQNITAIQL	169
Db	98	QGMAYIERKGHGKGRVAAHRVFERAAAOHQEARKAVAFSQMGFDSRWSIQEAKTV	157
QY	170	YESLAKEGSCKAQNALGFLSSYIGIME-YDQAKALIYYTFGSAGGNMWSOMILGYRLSG	228
Db	158	FEDLEKNGSPDQALGLFMHMGAGIGVEKSNQAKALVYWFYFSLGNGPLAQMAMGFYRSHG	217



```
QY 387 VPLNYAALYFQKAAKGWPDQAQFQGMYYSG-----SGIKWD----- 426
Db 353 PLFDDIDKAYWFKRGATKNDNSYGLGYMAYHGLTNGVDREKGMRLINLAVMNNPHA 412
QY 427 -----YKLAIFYFYLASOGOPALAIYYLAKWYATGTGVRSCRTAVELYKGV 473
Db 413 LMFGLGIRLEEARIEAYHLFRAATQKSVISYKYLADCYNGTGTSRSMISASLYK-- 470
QY 474 CELGHWAERFLTAIPAY-----KDGIDSSLYQXALLAEMGYEVAQSNFAFIL 521
Db 471 -----KFVEAIRASATSMALAEIDEYGFVHFSYFYLYAAQMGVALAEINAYILM 522
QY 522 ESKKANI-----LEKEMV-PMALLLNRAAIOGNFAFVKIGDIYHYGYGTRK 569
Db 523 DENKFLINSVFYFNTQSEQAADKFAFEYSRAAQQDIDAIFKLGDYIYYGIGTPK 582
QY 570 DYQTAATHYSIAANKYHNAQMFNLAYMEHGLGITKDTHLARLYDMAAQSPDAHIPV 629
Db 583 DYSKAYTCYKIAEQSSIGMGLWMAHMEYIGRQDDIYIARRLLD-ELSSNQNSYFPL 641
QY 630 LPAVMKLETHLLRLDIL 646
Db 642 KVAIFWINIHQLYIKLL 658

RESULT 4
T18454
hypothetical protein C0550w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18454
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18937
A:Accession: T18454
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-807 <LAW>
A:Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332547; PIDN:CAA15596.1
C:Genetics:
A:Map position: 3
A:Introns: 512/3; 638/1; 750/2
A:Note: C0550w

Query Match 11.0%; Score 391; DB 2; Length 807;
Best Local Similarity 22.1%; Pred. No. 1.1e-16;
Matches 152; Conservative 106; Mismatches 237; Indels 194; Gaps 20;

QY 24 EEHNKQKERNVTVQSVNEIK-QYLSHILEQRTSSNVINKRENLEKKKNQKIRIKG 81
Db 92 EEEKKKKLYLKNVDRANDLDSFSYIHNNEDSSNDNNGNKYNNRKNQNDNDNKG 151
QY 82 IQNKDIL--KRN----- 103
Db 152 YNNKQDNNKRNYSKONDKDYNNKYTYDDNNDYNNNEYNNDYNNIENKKEKEKEK 211
QY 104 TDEGDOLFQKMGKIVLOQSKOK-EEAYLLFAKADMGNLK-----AMEKMDALLFG-- 156
Db 212 ESRQKQFKLAIELKNGSKKKKNINKCIEIFOKLITDKNDKITSYSELGLIYFEGYK 271
QY 157 NFGV---QNTAAIQYLSIAKEGSCQAQNALGFLSSYGI----- 193
Db 272 NYFFSKRNVNLSLHYLQRAAMKNPAALHFLSFIFYDFPHKTEENKRNKKYIQNVQN 331
QY 194 -----GNEYDQ-----AKALIYYTSGAGNMMSQMLGYRSLGNVLONCEVAL 239
Db 332 VNNINNINYQNLOQNFNFKKSIEFEMIAASLNYPISLITLAYKFLYGNMKQNCYKAK 391
QY 240 SYKKVA-----DYIADTPEKSEGVPVEKVRLETERPENLSSNSEILDWDIYYKFLAE 293
Db 392 KLYKNVAENVMSDYI-----NIPLSELDLL-NGENLNMEINET----- 428
```

```
QY 294 RGDVQIQVSLGHLITGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAPVQ 353
Db 429 -----NMKNNEEIELEFLNEQIKGGDVMAMYDLGKKYE-----EK 465
QY 354 NNATAFKYSMAASKNAIQLHGLGLLYPHGKGVPLNYAEALKYFOKAAEKGWPDQAQFL 413
Db 466 NFKQAFKYINEASKNNLALRELGIYYLYGYGVQVDINKSIENFSAKAAEAGDVESKCYL 525
QY 414 GMYVSGSGIWKDYKLAIFY-----FYLA----- 437
Db 526 GTIYFIDG-YKNLEUSLYLIEAASHDYGAEFFFAEILDISMRKQVYISDYVYEVFK 584
QY 438 -----SQSQPLAIYYLAKWYATGTGVRSCRTAVELYKVCGLHWAERFLTAIFYK 492
Db 585 LYEHSDADLVGVQAYPREAQLYEIGKGVKQSCNLATLSYKFAESTLWINNIRQGM DYILE 644
QY 493 GDIDSSLYQXALLAEMGYEVAQSNFAFILESKKA-NILEKEKMYPMALLLNRAAIOGNA 551
Db 645 KDYLKAFYTYALASYEGYEIAQNLLVYIYRTNKLNNVHPKTI---MLVLNLLYKOGNY 700
QY 552 FARVKIGDYH-----YYGYGTRK 569
Db 701 KALYEMGEIYKEQNKKEELSVSYYKLGKK 729

RESULT 5
S48558
probable membrane protein YLR207w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L8167.5
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 06-Feb-1998
C:Accession: S48558
R:Pauley, A.
submitted to the EMBL Data Library, September 1994
A:Description: The sequence of S. cerevisiae cosmid 8167.
A:Reference number: S48545
A:Accession: S48558
A:Molecule type: DNA
A:Residues: 1-833 <PAU>
A:Cross-references: EMBL:U14913; NID:g544497; PID:g544502; MIPS:YLR207w
C:Genetics:
A:Gene: SGD:HRD3
A:Cross-references: SGD:S0004197; MIPS:YLR207w
A:Map position: 12R
C:Keywords: transmembrane protein
F:2-18/Domain: transmembrane #status predicted <TM1>
F:768-784/Domain: transmembrane #status predicted <TM2>

Query Match 10.4%; Score 373; DB 2; Length 833;
Best Local Similarity 23.9%; Pred. No. 1.5e-15;
Matches 165; Conservative 106; Mismatches 250; Indels 168; Gaps 28;

QY 97 KQAEKNF-----TDEGDOLFQKMGKIVLOQSKOKKEEAYLL-----FAKAA 138
Db 68 RNEEQYQSIQWNEITDSQRHYEL---LVQSEQFNNSSEATYTLQIHLWSQYFNPHNM 124
QY 139 DMGNLAKEMKADALLFGNFGVNITAAIQYLSIAKEGSCQAQNALGFLSSYIGMEYD 198
Db 125 TLAH-KYLEFNDLTHFTNH-----SAIFDLAVMTATGCCASGNDQTVIPQ-----D 170
QY 199 QAKALIYYTSGAGNMMSQMLGYRSLGNVLONCEVALSYKKYKVDYIADTFEKE- 257
Db 171 SAKALLYQRAAQGLNKAQVLAIFYSGFNVPNPFHKSILVLYRDIAEQLRKSYSDREW 230
QY 258 -----GVPVEKVR-LTERPENLSSNSEILDWDIYYK 289
Db 231 DIYFPYTESYNVRISDFESGLLGLKGLNSVFSVTVRKTRTPDIGS-----P 276
QY 290 FLAERGDVQIQVS---LGQLHLITGRKGL-----DQD-----YYKALHYFLKAAKA 331
Db 277 FIAQVNGVQMTLQIEPMGRFAFNQDNGNNGDEDEDAESERRIIRIYYAALNDY-KGTYS 335
```

Qy	332	GSAN-----AMAFIGKMYLEGNAAPQNNATAF	359
		: :     : :	
Db	336	QSRNCERAKNLELTQYKEFQPHVDNLDLPQVFFYVVRCLQLGHMVFYTGSGSKPNIHMAE	395
Qy	360	KYF--SMAASKGNAIGLHG-----LGLLYFHGKVPNLNVAEALKYFQKA----AEKGWPD	408
		:     : : : :     : :	
Db	396	EILTSLTSEISR-RAQGPPIGRACIDLGLI---NOYITNNISQAISYIMKAMTKTOANG--I	449
Qy	409	AQOFLGPMYSGSGIWKDYKLA--FKYFVYLASOGQPLAIYYLAKMYATGTGVRSCRTA	466
		:     : : : :     : :	
Db	450	VEFOLSKLATS---PPEEKIGDPNLMETAYLINGFIPIAYEFVMIESGMNKSSEVENT	505
Qy	467	VELYKGVCEIGH--WAKKFUTATYAKYDGDIDISLVOYALLAEMGYEVAQNSAFILBESK	524
		:     : : : :     : :	
Db	506	AYLFKTEVDKNEAIMAPKLTAFAALINORSEVALWAYSQLAEQGYETAQVGAAYLQYL	565
Qy	525	KANILE----KEKMPMALLWNRALTOGNAFARVKIDGYHYHYGYCTKKDYQTAATHYSI	580
		: : : : :     : :     : :     : :     : :	
Db	566	PYEFEDPRTTDORKTUAISYTRAFKQGNIDAGVAGDIYF-----OMQNYSKAMALYOG	621
Qy	581	AANKYHNAQMFNLAYMYEHGLGITTKDILHARLYDMAAQTPDAHIPVLFVAMKLEUTH	640
		: :     : :     : :     : :     : :     : :	
Db	622	AALY-SIQALWNLGYMHGHLGVNRDFHLAKRYIDOVSEHDHRFVLAESKLSVLKLV---H	677
Qy	641	LLRDILFFNOFTTRWNWLKDNTIGPHWD	669
		: : : : :     : :     : :     : :     : :	
Db	678	LKSWLTWITR--EKVYWKPKSSPLNPNED	704

```

RESULT      6
B64799      hypothetical protein b0644 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: B64799
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B64799
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-327 <BLAST>
A:Cross-references: GB:AE000169; GB:U00096; NID:q1786862; PIDN:AACT37345.1;
A:Experimental source: strain K-12, substrain MG1655

```

Query Match	8.8%	Score 314;	DB 2;	Length 327;
Best Local Similarity	29.8%	Pred. No. 2e-12;		
Matches 103;	Conservative	57;	Mismatches 134;	Indels 52;
				Gaps 10;

Qy	271	ENJSSNEI	DWDI	QYKYKFLA	ERGDIQ	VSIGUHL	IGRKGLDQ	-DYKALHYFLKAA	329
		:	:	:	:	:	:	:	
Db	11	DNLSID	-EITE-	-----	RAEKGD	EAQYIVG	-FYNRD	SAIDSPDEKAFYWLKAA	59
		:	:	:	:	:	:	:	
Qy	330	KAGSANAM	ATIGKMY	LEGNA	AVPQN	NATAFKY	PFMSA	ASKGAIGLHGLCLLYFHCKGVPL	389
		:	:	:	:	:	:	:	
Db	60	EOGCE	EAQY	SIGKY	TEDK	SRHKD	NEQAF	-WLKRAALOGHTFASNALG	116
		:	:	:	:	:	:	:	
Qy	390	NYAEAL	KYFOKAE	KWPD	QAQFOL	GPWY	YSGG	SIWKDYKLAIFYFLASQSGQPIAIYYL	449
		:	:	:	:	:	:	:	
Db	117	NYKEAV	WYQITAA	ESGMSY	AQNNL	GMWY	RNGNV	AKDYALAFWYKQAALOGHSDAQNNL	176
		:	:	:	:	:	:	:	
Qy	450	AKMYAT	GTGVV	RSCTA	VELYK	GVCE	LGHWA	EKFLTAYFAYKDGDISLSVGYALLAEMG	509
		:	:	:	:	:	:	:	
Db	177	ADLYED	CKGVAQN	-----	KTAAFW	-----	YLSA	QOQG	204
		:	:	:	:	:	:	:	
Qy	510	YEVAQS	ANSAF	ILES	KKAN	ILE	EKKM	TPMALLWNRAAIOGNAPARYKIDGYHYGYGPKK	569
		:	:	:	:	:	:	:	
Db	205	NRHAQ	FOI	WDY	NAGE	---	VDQD	TKQAMYLYLKAAQSGSYGVAYNTGYMYKHGGQVEK	260
		:	:	:	:	:	:	:	
Qy	570	DYQTA	ATHYS	IAANKY	HNQA	MENL	AYMVE	HGLGITKDIHLARLY	615

Db 261 DYQAAFEWFTKAA-ECNDATAWYNLAIMYHYGEGRPVDLRQALDLY 305

```

RESULT      7
T40051
hypothetical protein SPBC28F2.08c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40051
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Stevens, K.; Badcock, K.; Churcher
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21902
A:Accession: T40051
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-713 <MCD>
A:Cross-references: EMBL:AL121795; PIDN:CAB57937.1; GSPDB:GN00067; SPDB:SPBC28F2.08c
A:Experimental source: strain 972h-; cosmid c28F2
C:Genetics:
A:Gene: SPDB:SPBC28F2.08c
A:Map position: 2

```

Query Match	7.78;	Score	275.5;	DB	2;	Length	713;
Best Local Similarity	20.86;	Pred.	No. 1.4e-09;				
Matches	127;	Conservative	113;	Mismatches	193;	Indels	179;
Gaps	22;						

  

Qy	162	NITAAIO-----LVESLAKGSKCAQNALGFLSSYG----	IGMEYD	198
Db	48	NVSASIKTLPDLVTLTPFKLVKDPGYVEE-AEKGDPESQFLIAMLYAMGPDERLGISFP	106	
Qy	199	OAKAL--IYYTFGSAGNMMSOMILGYRYLSGNINVLNCEVALSYMYKKVADYIDATFEK-	255	
Db	107	RNEPLSRIFELSATQNTVTALLAYAKHLNGLSTPMSVDKGVELYKQVAHQHISCLVQPL	166	
Qy	256	-----SEGVPVEKVRTERPENLSSNSEILDWDIVQYKFLEAERG-DVOIQVISCOLH	307	
Db	167	SHPAPDIAAEYPVDLYDLS-----RTSSYSVQVKDDIVEYLKDYALRGNNISAHISLATIY	222	
Qy	308	LIGRGLDQDIYKALHYFLKAAKAGSANAMAFIGKMVLEGNAAVPONNATAFK-----	360	
Db	223	QYGTPGKLKDILAVKHVLAAILRL-----VNSGIPDSPSEAIKSIHNPR	267	
Qy	361	-----YFSMAASKGNAIGHG-----LGLLYFHCKGVPLNYACALKYFOKAA	402	
Db	268	HAPTTKETANSLSAARFLCGMALHGELGKPDPSLAYAWFE-YGVSLNHSSS-----KAA	321	
Qy	403	EKGWPAQFOQLGFGRMYSYSGSIGTWKYKIAFKFYFIASOSGOPLAITYYAK-----	451	
Db	322	-----IAYMFYCPVAENTESTITKLENALASNODPLAFAVACKVSLANGQIDEA	371	
Qy	452	----MYATGTGVRSCTRAVELYKG---VCBLGHWAKEF-----LTAY	487	
Db	372	TVHLIRAVNSGHLESVLHIADIIYYSNNQLSIAYI-ENFISRVLELDVKTISFDPLTRH	430	
Qy	488	FAYKDGDISDSLVQYALLAEMGYEVAOQSNAFIELESK---KANILEKEKMP-----	536	
Db	431	FAHR-----LSAELGNLMSQ-----ILAARDPSTSYLKITVIPTNEOTHNRN	473	
Qy	537	--MALLWNRAAIOGNFAFARVKIGDYHYGYGTTKDYQTAAATHYSIAANKYHNAQMNL	594	
Db	474	ARIAMNYSRAAARNHHISLIKIGDFYRMGLTSKTPELAFASYYSQAHAHPSALAYWRL	533	
Qy	595	AYMYEHGLGITPKDTHLARRLYDMAAQ-----TSPDAHIPVLFAVKL	636	
Db	534	GWMHEYGVGPVDEMAKKKNYDNALMHDTRAFTLARTLRMLRSSPDSFWSNIYRI LK	593	
Qy	637	ETHTLLRDILFF	648	
Db	594	VTYKFLKLVQYF	605	

```
RESULT 8
G71856
hypothetical protein jhp1045 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
C:Accession: G71856
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: G71856
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <ARN>
A:Cross-references: GB:AE001532; GB:AE001439; NID:g4155628; PIDN:AAD06623.1; PID:g415563
A:Experimental source: strain J99
C:Genetics:
C:Superfamily: conserved hypothetical secreted protein HP1098

Query Match 7.5%; Score 267.5; DB 2; Length 256;
Best Local Similarity 34.9%; Pred. No. 1.2e-09;
Matches 67; Conservative 32; Mismatches 88; Indels 5; Gaps 3;

QY 287 YKFLAERGDVOIQVSLGOLHILGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLE 346
DB 52 FYKSAIKNGESLAYILGIMYENGR-GVPKDYKKAVEYFQKAVDNDIPRGYNNLGVWYKE 110

QY 347 GNAAVPQNNTAFKYFMSAASKGNAIGLHGLLGYFHGKGVPLNYAEALKYFQKAEKGW 406
DB 111 GR-GVPKDEKKAVEYFRIATERYGTNAYINLGIMYMEGRGVPNSYAKATECFRKAMHKN 169

QY 407 PRAQFQFGFMYSGS---GIWKDYKLAFFYFLASOGQPLAIYYLAKMYATGTGVVRS 463
DB 170 VEAYILLGDIYSGNDQGLIEPKDKRAIYYKMAADVSSRAYEGLSESRYGLGVEKDK 229

QY 464 RTAVELYKGVC 475
DB 230 KKAEEYMQKACD 241

RESULT 9
E64659
conserved hypothetical secreted protein HP117 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 17-Mar-2000
C:Accession: E64659
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: E64659
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-256 <TOM>
A:Cross-references: GB:AE000518; GB:AE000511; NID:g2314268; PIDN:AAD08161.1; PID:g231426
C:Superfamily: conserved hypothetical secreted protein HP1098

Query Match 7.2%; Score 258.5; DB 2; Length 256;
Best Local Similarity 34.9%; Pred. No. 4.2e-09;
Matches 67; Conservative 29; Mismatches 91; Indels 5; Gaps 3;

QY 287 YKFLAERGDVOIQVSLGOLHILGRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLE 346
DB 52 FYKSAIKNGEPLAYVLGIMYENGR-GVPKDEKKAEEYFQKAVDNDIPRGYNNLGVWYKE 110
```

```
QY 347 GNAAVPQNNTAFKYFMSAASKGNAIGLHGLLGYFHGKGVPLNYAEALKYFQKAEKGW 406
DB 111 GR-GVPKDEKKAVEYFRIATERYGTNAYINLGIMYMEGRGVPNSYAKATECFRKAMHKN 169

QY 407 PRAQFQFGFMYSGS---GIWKDYKLAFFYFLASOGQPLAIYYLAKMYATGTGVVRS 463
DB 170 VEAYILLGDIYSGNDQGLIEPKDKRAIYYKMAADVSSRAYEGLSESRYGLGVEKDK 229

QY 464 RTAVELYKGVC 475
DB 230 KKAEEYMQKACD 241

RESULT 10
T46587
hypothetical protein [imported] - Vogesella indigofera
C:Species: Vogesella indigofera
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
C:Accession: T46587
R:van de Loo, F.J.; Keese, P.; Llewellyn, D.
submitted to the EMBL Data Library, September 1998
A:Description: Structural and regulatory genes controlling indigoidine production in
A:Reference number: Z23093
A:Accession: T46587
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-268 <VAN>
A:Cross-references: EMBL:AF088857; PIDN:AAD54009.1
A:Experimental source: ATCC 19706
```

```
Query Match 7.2%; Score 255.5; DB 2; Length 268;
Best Local Similarity 29.8%; Pred. No. 6.9e-09;
Matches 76; Conservative 40; Mismatches 102; Indels 37; Gaps 3;

QY 328 AKAGSANAMAFIGKMYLEGNAAPVONNATAFKYFMSAASKGNAIGLHGLLGYFHGKGV 387
DB 39 AAQSGDANAQFRIGVYERK-GSVPQDFAKATWYRQAATQGMPEAQNNLAVLYLNGQGV 97

QY 388 PLNYAEALKYFQKAEKGWPDQFQFGFMYSGGIWKDYKLAFFYFLASOGQPLAIY 447
DB 98 NQDAEALAWFRKAATQGGAEQALNGLMNGQTFPKNDDEAVVWTRKAAAGQLMADY 157

QY 448 YLAKMYATGTGVVRSCTAVELYKGYCELGHWAEKFLTAYFAYKDGDISSLYQYALLAE 507
DB 158 NLAIMMREGRLPQDDAAAVAFPRKV-----AE 185

QY 508 MGYEVAQSASFLESKKANILEKKNYPWALLLNRAIQGNARFARVKIGDYHYGYGT 567
DB 186 QGVIAQSNLGLMYKLGVRGAQD----YQLALSWLKRKGVAAQKSAQAQNLGVLYLEGKV 241

QY 568 KKDYQTAATHYSIAA 582
DB 242 AQDDNEAVVWFLAA 256
```

```
RESULT 11
G71958
hypothetical protein jhp0220 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: G71958
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: G71958
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <ARN>
A:Cross-references: GB:AE001460; GB:AE001439; NID:g4154733; PIDN:AAD05807.1; PID:g415
```







---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2002, 15:04:26 ; Search time 31.14 Seconds  
(without alignments)  
3236.404 Million cell updates/sec

Title: US-09-714-882-2  
Perfect score: 3570  
Sequence: 1 MKPLSLLEILILGVTKT.....LFVIGLIVGLILLRNHHG 689

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_invertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2650.5	74.2	505	4 Q9UGD3	Q9ugd3 homo sapien
2	1640	45.9	794	11 Q9ESM7	Q9esm7 mesocricetu
3	1264	35.4	819	5 Q9V415	Q9v415 drosophila
4	1159	32.5	685	5 Q20423	Q20423 caenorhabdi
5	1156	32.4	685	5 Q17378	Q17378 caenorhabdi
6	817	22.9	678	10 Q9LM25	Q9lm25 arabidopsis
7	723.5	20.3	604	10 Q9C6B6	Q9c6b6 arabidopsis
8	455	12.7	680	3 Q13875	Q13875 schizosacch
9	435	12.2	1029	4 Q94847	Q94847 homo sapien
10	435	12.2	1201	2 Q9XDH8	Q9xdh8 legionella
11	421.5	11.8	392	10 Q9FX49	Q9fx49 arabidopsis
12	391	11.0	807	5 Q77341	Q77341 plasmodium
13	373	10.4	833	3 Q05787	Q05787 saccharomyc
14	349	9.8	71	4 Q9NPY6	Q9npy6 homo sapien
15	294.5	8.2	469	2 Q9RN76	Q9rn76 coxiella bu
16	275.5	7.7	713	3 Q9USV0	Q9usv0 schizosacch
17	271.5	7.6	256	2 P94848	P94848 helicobacte
18	267.5	7.5	256	2 Q9ZK96	Q9zk96 helicobacte
19	258.5	7.2	256	2 Q25742	Q25742 helicobacte

20	255.5	7.2	268	2 Q9RBX2	Q9rbx2 pseudomonas
21	248	6.9	355	2 Q9ZMJ9	Q9zjm9 helicobacte
22	247	6.9	477	2 Q9JU20	Q9ju20 neisseria m
23	244	6.8	355	2 Q25021	Q25021 helicobacte
24	240	6.7	468	2 Q9JWX2	Q9jwx2 neisseria m
25	232	6.5	172	2 Q9CKK3	Q9ckk3 pasteurella
26	228	6.4	467	2 Q9JZ25	Q9jz25 neisseria m
27	226.5	6.3	290	2 Q9ZKB5	Q9zkb5 helicobacte
28	222.5	6.2	290	2 Q25728	Q25728 helicobacte
29	215.5	6.0	330	2 Q9A6P5	Q9a6p5 caulobacter
30	214	6.0	306	2 Q24968	Q24968 helicobacte
31	214	6.0	352	2 Q9ZLX8	Q9zlx8 helicobacte
32	212	5.9	198	2 Q9K0Y5	Q9k0y5 neisseria m
33	210.5	5.9	305	2 Q9ZMS0	Q9zms0 helicobacte
34	200.5	5.6	623	3 Q07622	Q07622 saccharomyc
35	199.5	5.6	242	2 Q9CNO9	Q9cng9 pasteurella
36	197	5.5	201	10 Q40097	Q40097 ipomoea tri
37	195.5	5.5	364	10 Q9S730	Q9s730 arabidopsis
38	195.5	5.5	392	2 Q66397	Q66397 acinetobact
39	194	5.4	95	4 Q12916	Q12916 homo sapien
40	189.5	5.3	687	3 P87065	P87065 candida alb
41	187	5.2	328	2 Q916D4	Q916d4 pseudomonas
42	183.5	5.1	251	2 Q9ZMA4	Q9zma4 helicobacte
43	183.5	5.1	512	3 Q94486	Q94486 schizosacch
44	183	5.1	510	11 Q9DCV6	Q9dcv6 mus musculu
45	176.5	4.9	225	2 Q25345	Q25345 helicobacte

## ALIGNMENTS

RESULT 1

Q9UGD3	PRELIMINARY;	PRT;	505 AA.
ID	Q9UGD3		
AC	Q9UGD3;		
DT	01-MAY-2000 (Tremblrel. 13, Created)		
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE	DJ842G6.2 (NOVEL PROTEIN IMILAR TO SEL1L (SEL-1 (SUPPRESSOR OF LIN-12, C-ELEGANS)-LIKE)) (FRAGMENT).		
DE	C-ELEGANS)-LIKE)) (FRAGMENT).		
GN	DJ842G6.2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Barlow K.;		
RL	Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; AL109657; CAB65792.2; -		
FT	NON_TER		
FT	NON_TER		
SQ	SEQUENCE 505 AA; 57003 MW; 536D4361FA826E35 CRC64;		

Query Match	74.2%;	Score	2650.5;	DB	4;	Length	505;
Best Local Similarity	99.8%;	Pred. No.	1.7e-163;				
Matches	505;	Conservative	0;	Mismatches	0;	Indels	1;
Gaps	1;						
QY	184	ALGFLSSYGIGMEYDQAKALIYYTFGSAGGNMMSQMLGYRLSGINVLQNCNEVALSYK	243				
DB	1	ALGFLSSYGIGMEYDQAKALIYYTFGSAGGNMMSQMLGYRLSGINVLQNCNEVALSYK	60				
QY	244	KVADYIADTFEKGVPVKEKRLTERPENLSSNSEILDWDIYQYKFLAERGVDQIQVSL	303				
DB	61	KVADYIADTFEKGVPVKEKRLTERPENLSSNSEILDWDIYQYKFLAERGVDQIQVSL	120				
QY	304	GQLHLIGRGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNAATAPKYS	363				
DB	121	GQLHLIGRGLDQDYKALHYFLKAAKAGSANAMAFIGKMYLEGNAAVPQNNAATAPKYS	180				
QY	364	MAASKGNATGLHGLGLLYFHGKGVPLNYAALKYFOKAAEKGWPDQAFQGLGFMYSGSGI	423				
DB	181	MAASKGNATGLHGLGLLYFHGKGVPLNYAALKYFOKAAEKGWPDQAFQGLGFMYSGSGI	240				

```
QY 424 WKDYKLAFFYFLASQSGOPLAIIYLLAKMYATGTGVVRSCTAVELYKGYCELGHWAERF 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 WKDYKLAFFYFLASQSGOPLAIIYLLAKMYATGTGVVRSCTAVELYKGYCELGHWAERF 300
QY 484 LTAIFYAYKGGIDSSLVQVALLAEMGYEVAQNSAIFLESKKANILEKEKMYPMALLLN 543
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 LTAIFYAYKGGIDSSLVQVALLAEMGYEVAQNSAIFLESKKANILEKEKMYPMALLLN 360
QY 544 RAAIOGNAFARVKGIDYHYGYGTTKDYOTAATHYSTAANKYHNAQAMFNLAEMYEHGLG 603
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RAAIOGNAFARVKGIDYHYGYGTTKDYOTAATHYSTAANKYHNAQAMFNLAEMYEHGLG 420
QY 604 ITRDIHLARLYDMAAQTSPDAHIPVLFVAVMKLETTTHLLRDILFFNQFTTRWNWKLKDN 663
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 ITRDIHLARLYDMAAQTSPDAHIPVLFVAVMKLETTTHLLRDILFFN-FTRWNWKLKDN 479
QY 664 IGPBWDLFVIGLIVPGLILLRNHHG 689
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 480 IGPBWDLFVIGLIVPGLILLRNHHG 505

RESULT 2
Q9ESM7 PRELIMINARY; PRT; 794 AA.
AC Q9ESM7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SELIL.
GN SELIL.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
SEQUENCE FROM N.A.
RA Wada M., Moriyama T.;
RT "Hamster kidney SELIL.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB048195; BAB12403.1;
DR InterPro; IPR000562; FN_Type_II.
DR Pfam; PF00040; fn2.1.
DR PRINTS; PD00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II.1.
DR SMART; SM00059; FN2.1.
DR PROSITE; PS00023; FIBROTECTIN_2; UNKNOWN_1.
SQ SEQUENCE 794 AA; 88534 MW; 610F6930114FE2CC CRC64;

Query Match 45.9%; Score 1640; DB 11; Length 794;
Best Local Similarity 52.2%; Pred. No. 7.3e-98;
Matches 314; Conservative 112; Mismatches 155; Indels 20; Gaps 4;

QY 99 AKENFTDEGDLFKMGIKVLOQSKSQKQKEAYLLFAKAADMGNLKAEMKADALLFGNF 158
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 175 AKRQMQEAMIIYQAGKTLINGSNRKSQREAYRYLQKAAGMHTKALERSVALLFGDY 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 159 GYONITAAIOLVESLAKESCAONALGFLSSVIGMEYDQAKALIYYTSGSAGGNMSQ 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 LTQNIQAEMEFKLTGEGSPKGOTALGFLYVSGLVGNSQAKALIYYTFGALGGNLIH 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 219 MILGYRLSGINVLQCEVALSYKKVADYIADTFEKSQGVPEKVRLETERPENLSSNE 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 295 MVLYGYRWAGVLOSCSVLTHRYLVANHVASDISLTGSGVVQRIRLPDEVENPGWNSG 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 279 ILWDNDIYQYKFLAERGDVQIVSQGLHIGKGLDQDYQYALHYFLKAAKAGSANAMA 338
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 355 MLEEDIYIYQFLAERGDVQIVSQGLHIGKGLDQDYQYALHYFLKAAKAGSHAMA 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 339 FTGKMYLEGNAAVPONNATAFYFESMAASKGNAIGHLGLGLYFHGKGYPLNVAEALKYF 398
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 415 FLGKMYSEGSDIVPQSNETALHYFKKAADMGNPVGQSLGMAYLYGRGIQVNDLALKYF 474
QY 399 QKAAEKGWDAQFOLGFMYSYSGGIWKDYKLAFFYFLASQSGOPLAIIYLLAKMYATGTG 458
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 QKAAEKGWDAQFOLGFMYSYSGGIWKDYKLAFFYFLASQSGOPLAIIYLLAKMYATGTG 534
QY 459 VVRSCTAVELYKGYCELGHWAERFVAYKGGIDSSLVQVALLAEMGYEVAQNSA 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 VVRSCTAVELYKGYCELGHWAERFVAYKGGIDSSLVQVALLAEMGYEVAQNSA 594
QY 519 FILESKKANILEKEKMYPMALLLNRAAIOGNFAFARKYIDGYHYGYGTTKDYOTAATHY 578
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 FILDQREATIVGNETYPRALLHWNRAASQGYTVARIKLGYDFHFGTGDVDTAFIHY 654
QY 579 STAANKYHNAQAMFNLAEMYEHGLGTTKDIHLARLYDMAAQTSPDAHIPVLFVAVMKLE 638
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 RLASEQOHSQAAMFNLCYMKHEKGLGKQDIHLAKRYDMAAEASPDQVDFVFLACKLG 714
QY 639 THLL-----RDILFFNQFTTRWNWKLKDNTPGHWDLF---VIGLIVPGLILLRNH 687
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 VYFLQVIREANIRDI--FTQ-----LMDQLLGPWDILMTIALLIGTVIAYRQRQ 765
QY 688 H 688
Db 766 H 766

RESULT 3
Q9V415 PRELIMINARY; PRT; 819 AA.
AC Q9V415;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE BCDNA:LD23587 PROTEIN.
GN BCDNA:LD23587 OR CG10221.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballaw R.F., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng C., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houlton D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```



```
QY 522 ESKANIL---EKKMYPMALLNNRAALQGNAPARVIGDYHYGYGKDKDYQTAATHY 578
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 518 DRGEATSFSGPKDNMNERAFLNWORSANQYAAARVKGIDYIYGLGTEVDHSLAFSNY 577
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 579 STAAKYNHQAAMFNLAYMYEHGLGITKDILHARLYDMAAQTSPDAHIPVLFAVMKLET 638
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 578 KMAVDXHGVAQAMFNLGYMHVEGEGITRDLYLAKRFDQAIEHSQDAYMPSKLALAKLAF 637
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 639 THLLRDILFFNQFTTRWNWKL-----DNMTGPHWD--LFVIGLIVPGLILLRN 686
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 VFYLEEL-----NKLPLISPMKTVGPRWDAILMTVSALVP-LFLFWRH 680
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
Q17378 PRELIMINARY; PRT; 685 AA.
AC Q17378;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE SUPPRESSOR/ENHANCER OF LIN-12 (SEL-1).
GN SEL-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=96304591; PubMed=8722778;
RA Grant B., Greenwald I.;
RT "The Caenorhabditis elegans sel-1 gene, a negative regulator of lin-12
RT and glp-1 encodes a predicted extracellular protein.";
RL Genetics 143:237-247(1996).
DR EMBL; U50828; AAC47112.1; -.
SQ SEQUENCE 685 AA; 76222 MW; 3779A39937A4FC52 CRC64;
```

```
Query Match 32.4%; Score 1156; DB 5; Length 685;
Best Local Similarity 38.7%; Pred. No. 1.2e-66;
Matches 253; Conservative 136; Mismatches 215; Indels 50; Gaps 11;

QY 53 EORTSSNVINKRENLEKKKQKIRIKQIONKDIK-RNKNHLOKQAEKNFTDEGDQLF 111
      || || || | | | | | | | | | | | | | | | | | | | | | | | |
Db 57 EQVTSQDENK-----SNREIPKVISEYLAEKVEQPPSPAEAEF 97
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 112 KMGIVLQSK--SOKQEEAYLLFAKADMGNLKAMEKMDALLFGNGVQNTAAIQL 169
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 QRCMAYIERGKGHGRVAAHRVFERAAAOHQBARAVAFSOMFGDYSRWSIOEAKTV 157
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 YESLAKEGCKAQNALGFLSSYGIME-YDQAKALIYKTFGSAGGNMMSOMILGYRILSG 228
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 158 FDELEKNSPDQLAGFMHMGAGIVEKSNQAKALVITFSAUGNPLAQMANGFRYSHG 217
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 229 INVQNCVALSYKKVADYIADTFEKGSEGVPEKVRITERPE---NLSSNSEILDWIIY 285
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 218 VGPVQNCETALSYYQKAVTVVDNVKFTTGQIQLRLTDETDPTIHQPGSAPLESNLL 277
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 286 QYGFLEAGRDVQIQLGOLHIGKGLDQDYKALHYFLKAAKAGSANAMAFIKMYL 345
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 278 EYKMLADKGDTSAGLGLQIYLAGRGLNQNFELAFRYLLAAAECSADALTYLGKMYL 337
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 346 EGNAAVPPQNATARKYFESMAASKGNAIGLHGLGLLYFHCKGVPLNAYEALKYFQKAAEK 405
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 338 DGTPTPKDYQKSEYLMKSDKSPSAQAVLGMKGVKKKNYKALKLLTSLADKK 397
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 406 WPDQFQGLGFMFYSG-----SGIWKDYKIAFYFLASQGPPLAIYILAKMYATGTGYVR 461
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 NADGOMYLAELHYKVPINKGVHRDFKSKVLYQLASQNGHILAYNLQAQHAAGTGVPR 457
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 462 SCRTAVELYKGVCELGHWAEKFLTAIFYAYKKGIDSSLVQYALLAEMGYEVAQNSAFTL 521
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 458 SCSHAVDLFKSVAERKGRMEAHSAISKONRVDEAAAMKYLFAELGYEVAQTNLAYIL 517
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 522 ESKANIL---EKKMYPMALLNNRAALQGNAPARVIGDYHYGYGKDKDYQTAATHY 578
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 518 DRGEATSFSGPKDNMNERAFLNWORSANQYAAARVKGIDYIYGLGTEVDHSLAFSNY 577
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 579 STAAKYNHQAAMFNLAYMYEHGLGITKDILHARLYDMAAQTSPDAHIPVLFAVMKLET 638
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 578 KMAVDXHGVAQAMFNLGYMHVEGEGITRDLYLAKRFDQAIEHSQDAYMPSKLALAKLAF 637
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 639 THLLRDILFFNQFTTRWNWKL-----DNMTGPHWD--LFVIGLIVPGLILLRN 686
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 638 VFYLEEL-----NKLPLISPMKTVGPRWDAILMTVSALVP-LFLFWRH 680
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
Q9LM25 PRELIMINARY; PRT; 678 AA.
AC Q9LM25;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE T10022.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chioi J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T10022 from chromosome
RT I.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC069551; AAF78381.1; -.
SQ SEQUENCE 678 AA; 75941 MW; 415235B57FC7430D CRC64;
```

```
Query Match 22.9%; Score 817; DB 10; Length 678;
Best Local Similarity 36.6%; Pred. No. 9.6e-45;
Matches 189; Conservative 99; Mismatches 185; Indels 44; Gaps 12;

QY 137 AADMGNLKAMEKMDALLFGNGVQNTAAIQLYESLAKEGCKAQNALGFLSSYGIM- 195
      || || || | | | | | | | | | | | | | | | | | | | | | | | |
Db 99 AASEGFRMLEAEVDEI-----EASSAGDPHAQSGIMGFV--YGIGMM 139
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 196 -EYDQAKALIYTFGSAGGNMMSOMILGYRILSGINVLQNC-EVALSYKKVADYIADTF 253
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 140 REKSKSFLHNFHFAAGNMOSKMAFYTLR-----ODMHDKAVQYAEAEAVNSF 194
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 254 EKSEGV-VEKVRV---TERPENLSSNSEILDWIIYQYKFLAERGDVQIQLGOLHIL 309
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 LISKDSPVVEPRIHSGTEENKALRKSGEDEDFQILEYQAKGNANAMWKIGLYFF 254
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 310 GRKGLDQDYKALHYFLKAAKAGSANAMAFIKMYLEGNAAPVQNNATAFKYSMAASKG 369
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 GLGRERDHTKALHFLKAVDKGSEPMELGELIYARG-AGVERNTRYALEMLTIAAREG 313
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 370 NAIGHLGLGLLYFHCKGV-PLNAYEALKYFQKAAEKGFDPQFQGLGFMYSYSGIWKDYK 428
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 314 LYSAPNGYLYGVKGVYDKNYTKAREYFEKAVONEDPSGHYNGLVLYLKIGVNRDVR 373
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 429 LAFKYFYLASQSGQPLAIYILAKMYATGTGVVRSCTRAVELYKGVCELGHWAEKFLTA 488
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 374 QATKFFVVAANAGQPKAFYQLAKMFTGVGLKKNLEMATSFYKLVARCPNWSLSRWALE 433
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 489 AYKDGIDSSLVQYALLAEMGYEVAQNSAFILSKKANIL-----EKKMYPMAL 539
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

Db 434 AYLGKGVGRKALILYSMAEMGEVAQSNAAWLLDKYGRSMCMGVSGFCTDKER-HERAH 492
QY 540 LLNWRRAAIOGNARFARVKIGDYHYGYGKDYQTAATHYSIAANKYHNAQAAMFNLAYWE 599
Db 493 SLWRASEQNEHAALLIGDAYIYGRGTERDFVRAEAY-MHAKSOSNAQAAMFNLYMHE 551
QY 600 HGLGITKDIHLARRLYDMAAQTSPDAHIPVLFAVMKL 636
Db 552 HQGLPFDLHLAKRYIDYDESLSQDAARLPVTALASL 588

RESULT 7
Q9C6B6 PRELIMINARY; PRT; 604 AA.
AC Q9C6B6
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 67.8 KDA PROTEIN.
GN F6D5.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altar J., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu X.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Miltscher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC079676; RAG51809.1; -.
KW Hypothetical protein.
SQ SEQUENCE 604 AA; 67761 MW; 973AFD4E38D7EBA6 CRC64;

Query Match 20.3%; Score 723.5; DB 10; Length 604;
Best Local Similarity 34.4%; Pred. No. 9e-39;
Matches 177; Conservative 90; Mismatches 160; Indels 87; Gaps 12;

QY 137 AADMGNLKAEMKADALLFGNFGVQNTAAIQLYESIAKEGSKAGNALGFLSSYGIGM- 195
Db 100 AASEGNTLME-----AVSEIDSSAGSDPHQAQSVMGFV--YIGIM 140

QY 196 -EYDQAKALIYTFGAGNMMSOMILGYRYSGLNVLQNC-EVALSYKKVADYIADTF 253
Db 141 RETSRKSLIHFFHAAGNMOSKMAAFRYLR-----QNMVDKAVELAEIAETAVNSF 195

QY 254 EKSEGVV-EKVL-----TERPENSSNSELTDWDIYQYKFLAERGDVQIQVSLQHLI 309
Db 196 LISKDSPMAEPVRIHTGTEENKDALRSGEDEDQILEYQAEKNSVAMHKIGLYFYF 255

QY 310 GRKGLDQDYKALHYFLKAAKAGSANAMAFIGKMWLEGNAANPONNATAFKYFSMAASKG 369
Db 256 GURGLRRDHAKALYWFSKAE----- 275
```

```

QY 370 NAIGHGLGLLYFHGKV-PLNYAEALKYFQKAAEKGPDAQFQLGFMYYSGSGIWKDYK 428
Db 276 ---ENGLGYLVKGVGDKRNYTKAREYFEMAANNEDPSGHYNLGVLVYKGTGVKQDVR 331
QY 429 LAFKYFYLASOGQPLAIYYLAKMTATGTGVVRSCTVELYKGVCELGHNAKEFLTAYF 488
Db 332 HATKYFFVAANAGOPKAFYQLAKMFHTVGLTKNEMATTFYKLVARGPWSLSRWALE 391
QY 489 AYKGDGIDSSLVQYALLAEMGEVAQSNASFLESKANIL-----EKEKMPMAL 539
Db 392 AYLGKGVGRKALILYSMAEMGEVAQSNAAWLLDKYGRSMCMGVSGFCTDKER-HDRAH 450
QY 540 LLNWRRAAIOGNARFARVKIGDYHYGYGKDYQTAATHYSIAANKYHNAQAAMFNLAYWE 599
Db 451 SLWRASEQNEHAALLIGDAYIYGRGTERDFVRAEAYMYAKSQ-SNAQAAMFNLYMHE 509
QY 600 HGLGITKDIHLARRLYDMAAQTSPDAHIPVLFAV 633
Db 510 HQGLPFDLHLAKRYIDYDALQSDTAAKLPVTAL 543

RESULT 8
O13875 PRELIMINARY; PRT; 680 AA.
AC O13875
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 78.7 KDA PROTEIN C1B3.10C IN CHROMOSOME I PRECURSOR.
GN SPAC1B3.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO C.ELEGANS SEL-1.
DR EMBL; Z98598; CAB11247.2; -.
KW Hypothetical protein; Signal.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 680 HYPOTHETICAL PROTEIN C1B3.10C.
FT DOMAIN 573 576 POLY-TYR.
SQ SEQUENCE 680 AA; 78703 MW; 522BD933D91FEC9 CRC64;

Query Match 12.7%; Score 455; DB 3; Length 680;
Best Local Similarity 26.3%; Pred. No. 2.5e-21;
Matches 162; Conservative 97; Mismatches 218; Indels 140; Gaps 21;

QY 126 QKEEAYLFLFAKADMGNLKAMEKMDALLFGNFG-VQNTAAIQLYESIAK-EGSKAQN 183
Db 86 QRNHATELLRSASHDNTDAMLYLANIEFFGLFEIPEIEDSVKYYDMLQKANGSAFANN 145

QY 184 ALGFLSSYIGMEY---DOAKALIYTFGAGNMMSOMILGYRYSGLNVLQNC-EVALS 240
Db 146 MNGFTYSTFS-EYASNNPALARIHWEAAKQSGDADHQLAYHNLIATLNMPSQSEAVK 204

QY 241 YKKVADYIADTFEKSEGVPEKVRLETERPENSSNSEILD--WDIYQYKFLAERGDVQ 298
Db 205 HYKFTSDHL---FE-----EECGSNVTYLCIWPFIQDYNAGENG--- 242

QY 299 IQVSLGQLHLIGRGLDQDY-----YKALH-----YFLK-----AAKAGSANAMAFIGKMWLE 346
Db 243 -----MGVGAASATYSDAYQALHTRSQYLREMSNSTEDWDYELMPEVAKLRH 292

QY 347 GNAANVPONNATA-----FKYFS-----MAASKGNAIGHG-----LGLLYFHGKG 386
Db 293 GMYKPRNRYTSDVLFKRVSRQYWPYTSNSVLANTPQSIILAAQSCGYLGLLHFLDKG 352
```

QY	387	VPLNYEALKYFOKAAEKWDPDAQOLFGEWYSYG-----SGIWKD-----	426
		: :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db	353	PLFDIDKAYMWFKRGAATDNSNYSYGLGYMAYHGLTNGVDREKMRLLINLAVMNPHA	412
QY	427	-----YKLAFKYFYLASOSGOPLATYYLAKMYATGTGVVRSRCAVELEYKGV	473
Db	413	LMFGLGLRLBEAREYEAYHLFLRAAATQSVISVKYLADCYNGTGTSRSMISASLYKK--	470
QY	474	CELGHWAEKELTAYFAY-----KDGDISLSVOYALLABMGVEVAQSNSATIL	521
Db	471	-----KFVEAIRASATSMTALEBIDEYGYFHNSFVYIIYAQAOMGYALAEINAATIM	522
QY	522	ESKKANI-----LEJEXMY-PMALLLNWRAAIQCNARFARKIGDYHYVGYGTYKK	569
Db	523	DENKEFLINSVFYFNVTQSEQAHHDKFAFEFSRAAAQGDDIAIFKLGDYYYYYGTPK	582
QY	570	DYOTAAATHSYIAANKYHNQAOMENLAYMEYHEGLGITKDHLARRLYDMAAQTSPOAHIPV	629
Db	583	DYSKATCYKIAYEQSSIGMLWNMAIMEYIGRDQDIYIARRLLJD-ELSSNQNSYEPFL	641
QY	630	LFAVMKLETTHLRDLIL	646
Db	642	KVAIFWINIHOLYIKLL	658

[illegible]

Qy	327	KAAGKAGSANAMAF	IGKMYLEGNAAVQNNNA	TAFKYFSMAASKGNATGLHGLGLLYPHGKG	388
Db	616	KGALTEDEP	AIYDIAIVLFKGGVKKRRRLA	LELMKAASKGLHQAVNGLGWYHKPK-	674
Qy	387	VPLNYAALYFQKAAEK	GHDPDAQFQGLGFMYSG--	-SGI-WKDYKLAIFYFYLASOSGOP	443
Db	675	--KNYAKAAYWLKAE	EMGNPDASNYLVLHLDGI	FFGVPGRNOTLAGIFYFKAAGGHH	732
Qy	444	LAIYVLAKMYATGT-	-GWRASCR	TAVELYKGVC-----LCHNAEKLTFAYFAYKGGDIDS	497
Db	733	EGTLWCSLIYITGNL	TFPRPEKAVVWAKHVAE	KNGYLCHVIRKGLNAYL---EGSWHE	789
Qy	498	SLVOYALLAEMGEY	VAOSNAFLESKANILEK	KWYPMAILLW---NRAAIQNA--F	552
Db	790	ALLYVLLAAETGIE	VSOTNLAHICEERP----	DLARRYLGVCVWRYYNFSVFQIDAPSF	845
Qy	553	ARVIGIDVHYVGY-	GTKKDYQTATHYSTIA	KNYHNAQAMFNLAYMYEHLGT-----	604
Db	846	AYLWMDLIYIGHON	QSDLSVQMYAQALD-	GOSQGFNLALIEDGTTIPPHILDF	904
Qy	605	-----TKDIHLAR	LYDMA-AQTS	PDAPHVLFAVMKLETHLLRIL	646
Db	905	LEIDSTLHNNISIL	QELYERCSHNSNEES	FSPCSLAWLYLHLRLWGAIL	955
RESULT 10					
Q9XND8	ID	PRELIMINARY; PRT; 1201 AA.			
AC	O9XND8	ID			
DT	01-NOV-1999	(T=EMBLrel. 12, Created)			
DT	01-NOV-1999	(T=EMBLrel. 12, Last sequence update)			
DT	01-JUN-2001	(T=EMBLrel. 17, Last annotation update)			
DE	ENHANCED ENTRY PROTEIN ENHC.				
GN	ENHC				
OS	Legionella pneumophila.				
OC	Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;				
OC	Legionellaceae; Legionella.				
OX	NCBI_TaxID=446;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	STRAIN=130B;				
RA	Cirillo S.L.G., Gupta M., Lum J., Cirillo J.D.;				
RT	"Isolation of Novel Loci Involved in Entry by Legionella pneumophila.";				
RL	Submitted (APR-1998) to the EMBL/GenBank/DBSJ databases.				
DR	EMBL; AF057704; AAD41587.1; -				
DR	InterPro; IPR001440; TPR.				
SQ	SEQUENCE 1201 AA; 134365 MW; 5C4ABEE5F3526DC4 CRC64;				
Query Match 12.2%; Score 435; DB 2; Length 1201;					
Best Local Similarity 24.2%; Pred. No. 1.1e-19;					
Matches 183; Conservative 102; Mismatches 246; Indels 224; Gaps					
Qy	34	NVTQSVNEIKOYLSH	LEQRTSSNVINKRENLL	EKKKQKIRIKGTQNKDILKRKN	93
Db	498	NAQTFEIGOLFQY	GTGLMDDASATIF--	YENAEQKHLGAEYNL-GTL---YLKRGKD	551
Qy	94	HLQQAENKFTDE	GDOLF-----	MGIKVLQOS-----KSKQKEAEVLLFAK	136
Db	552	ENDYQQALNWLT--	DSAFKGNKRAQYVLAR	LROGIVGPDGKEYIKANDEQAMMLYLSA	609
Qy	137	AADMGNLK-----	-----AMEKMDALL-----FGNF- 158		
Db	610	ANDYGPAYELAEY	LARDYNGLSVDRKQK	IALIRKLYQGAVKNKVAEALLPLAFYNAM	669
Qy	159	--GVQNTAIALQY	ESLAKEGSCKAQNALG	FLSSYIGMEYDQAKALIYYTFSAGGNWM	216
Db	670	DDNRKQEQAFKVA	EQAETGNEAALLGLM	LYDRGIGITADPAKAWYQ--QAGQNPV	727
Qy	217	SOMILGYRYSLG	INLQNCVALSYKKVA-----	-----DYIADTFEKS	256



[illegible]

```
Db 429 -----NMKNNEEIELEFLNEQIKGDMVAMVDLGRKKYK-----BK 465
QY 354 NNATAFKYSMAASKGNAIGLGLLPHGKGVPLNVAEALKYFOKAAEKGWPDAAQFOL 413
Db 466 NFOKAFKYNINASKNNLALALGLIYLYGVGVQKDKINKSTENFSAADGVSCKYLY 525
QY 414 GPMYSGSGIKWDYKIAFKY-----FYLA-----437
Db 526 GYIYFIDG-YKNLESLKLYLEAASHDYGEAFFFLAEIILDISMRKQYISDYVYEVVFK 584
QY 438 -----SQSGOPLAIYVLAKMYATGTVWRSCTAVELYKGCVELGHWAEKFLTATPAYKD 492
Db 585 LYEHSDALGVQVAFREAOLEYIGKGVKQSCNLATLSYKFIAESTLWINNRQGM DYLE 644
QY 493 GDIDSSLVQYALAEAGYVVAOSNFAILESCKA-NILEKEKMPWALLLNRAAIOGNA 551
Db 645 KYDKAFYFALASYEGYEIAQNLNLYIYRTNKLNNVHPRKI-----MLVNLLYKQGN 700
QY 552 FARVRIGDYH-----YGYGTTK 569
Db 701 KALYENGEIYKEONKEELSYYKGLKK 729

RESULT 13
Q05787
ID Q05787 PRELIMINARY; PRT; 833 AA.
AC Q05787;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CHROMOSOME XII COSMID 8167.
GN HRD3 OR L8167.5 OR YLR207W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=9432;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
RA Entlan K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Klotz K., Kottler P.,
RA Louis E.J., Messenguy F., Mewes H.W., Miosga T., Mostl D.,
RA Muller-Auer S., Wentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Pauley A.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Waterston R.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C (AB972);
RA Cherry J.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U14913; AAB67427.1; -.
DR SGD; S0004197; HRD3.
SQ SEQUENCE 833 AA; 95480 MW; E7DB29FBA16D9BFD CRC64;
```

```
Query Match 10.4%; Score 373; DB 3; Length 833;
Best Local Similarity 23.9%; Pred. No. 6.7e-16;
Matches 165; Conservative 106; Mismatches 250; Indels 168; Gaps 28;

QY 97 KOAENF-----TDEGDLQFKMGIKVLOQSKSKQKEEAYLL-----FAKAA 138
Db 68 RNEEKYQSIWONEITDSORHIEL---LVQSEQFNNEATYTLQIHLWSQYNFPHNM 124
QY 139 DMGNLKAMEKMDALLFGNFGVQNTAAQLYELSAKESCAQNALGFLSSYGYGMEYD 198
Db 125 TLAH-KYLEKFNLDLTHFTNH-----SAIFDLAVMTATGCCASGNDQTVIPO-----D 170
QY 199 QAKALYITFFSAGGMMMSQMLGYRLSGINVLQCEVALSYKVKVADYIADTFEKS- 257
Db 171 SAKALLYQRAAQLGNLAKAKQVLAYKYSGFVPRNFHSLVLYRDIAQLRKSYSRDEM 230
QY 258 -----GVPVEKVR-LTERPENLSSNSEILDWDIYYK 289
Db 231 DIVFPYWESYNNRISDFESGLLGKGLNSVPSTVRKTRTRPDIGS-----P 276
QY 290 FLAERCDVQIOVS---LQOLHLIGRKL-----DOD-----YKALHYFLKAAKA 331
Db 277 FIAQVNGVOMTLOIEPMGRFAPFNGDNGINGDEDEDASERRIIRIYYAALNDY-KGTYS 335
QY 332 GSAN-----AMAFIGKMYLEGNAAPQNNATAF 359
Db 336 QSRNCERAKNLELTYKEFOPHVDNLDPLQVYVYVRCLOLLGHMYFTGESSKPNHMAE 395
QY 360 KYF--SMAASKGNAIGLHG-----LGLLYFHGKGVPLNVAEALKYFOKA-----AEKGWPD 408
Db 396 EILTTSLERS-RAQGPPIGRACIDLGLI---NOYITNNISQISYMKAMKQANNG--I 449
QY 409 AQOLFQMYSSGSIWKDYKLA--PKYFVLASQSGOPLAIYVLAKMYATGTVVRSCTA 466
Db 450 VEFQSKLATS---FPEKIGDPFNLMTAYLNGFIPAIYEFVAMIESGMKSKSVVENT 505
QY 467 VELYKVCCELGH--WAEKFLTAYFAYKGDIDSSLVQYALLAEEMGYEVAQSAFTLESK 524
Db 506 AYLEKTFVDKNEAIMAPKLRTAFALINDRSEVALWAYSQLAEQGYETAQVSAAYLMYOL 565
QY 525 KANILE-----KEKMPMALLLNWRAIQGNARFARVIGDYHYGYTKDYQTAATHYSI 580
Db 566 PYEFEDPPRTDQRTKTLAISYTRAFKQGNIDAGVAGDIYF---OMONYSKAMALYQ 621
QY 581 AANKYHNAQAMENLAMYEHGLGITKDIHLARLDYMAQTSFDAHIPLVLFVAMKLETH 640
Db 622 AALKY-SQAIWNLTGMHEHGLGVNRDHLARYIDQVSEHDHRYFLASKLSVYKLD---H 677
QY 641 LLRDILFFNQFTRWNWLKLDNTIGPHWD 669
Db 678 LKSWLTWITR--EKVNYMKPSSPLNPED 704

RESULT 14
Q9NPY6
ID Q9NPY6 PRELIMINARY; PRT; 71 AA.
AC Q9NPY6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE DJ631M13.1 (SIMILAR TO MOUSE SELIL (SEL-1 (SUPPRESSOR OF LIN-12,
DE C.ELEGANS)-LIKE.) (FRAGMENT).
DN DJ631M13.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Laird G.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117333; CAC01943.1; -.
```

FT	NON_TER	71	71
SQ	SEQUENCE	71 AA;	7756 MW; 3EF2C63469879DF4 CRC64;

Query Match	9.83;	Score	349;	DB	4;	Length	71;
Best Local Similarity	100.0%;	Pred. No.	6.5e-16;				
Matches	71;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;

QY	113	MGIKVLQOSKQKEEAYLLFAKAADMGNLKAMEKMAADALLEGNGFVQVNTAAIQLYES	172
Db	1	MGIKVLQOSKQKEEAYLLFAKAADMGNLKAMEKMAADALLEGNGFVQVNTAAIQLYES	60
QY	173	LAKEGSCKAQN	183
Db	61	LAKEGSCKAQN	71

RESULT	15	
Q9RN76	PRELIMINARY;	PRT; 469 AA.
Q9RN76		
AC	Q9RN76;	
DT	01-MAY-2000 (TremBLrel. 13, Created)	
DT	01-MAY-2000 (TremBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TremBLrel. 13, Last annotation update)	
DE	IMMUNOREACTIVE PROTEIN.	
OS	Coxiella burnetii.	
OC	Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;	
OC	Coxiella group; Coxiella.	
OX	NCBI_Taxid=777;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-NINE MILE PHASE I;	
RA	Zhang G.Q., Hota A., Yamaguchi T., Fukushi T., Hirai K.;	
RT	"Cloning and characterization of the Coxiella burnetii gene encoding an immunoreactive protein."	
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AF185288; AAF01236.1; "	
SR	SEQUENCE 469 AA; 52695 MW; DDB3341C820011D5 CRC64;	
DQ		

Query Match	8.28;	Score	294.5;	DB	2;	Length	469;		
Best Local Similarity	24.34;	Pred.	No. 3.4e-11;						
Matches	132;	Conservative	86;	Mismatches	190;	Indels	135;	Gaps	22;

  

Qy	47	YLSHILEORTSSNVINKRENLEKKKKQRIKRIQIONKDI-----LKRNNKHLOKQAE	100
Db	21	YVSGFL--NPDNVYSLETQNNKKKAYALQKAASGLEKADVALGYFYLOQNOTTL--AE	75
Qy	101	KNFTDEGDLFKMGIKVYLOQSKSKQKEEAYILLPAKADMG-----NUKAMEKMA	151
Db	76	KTFE-----IAQRAYTNDPEAAMLLAILDRGFGVNRNRSRKSAEILERLSK	122
Qy	152	-----ALLFCNFGVQN---ITAAIQIYESLAKESCKAQNALGFLSSYIGCMEYDOAKA	202
Db	123	QNNAIQAQFMLNGYILKNKRKENIAISLLEKSANQNGYAKYNLAILAK-----QNK-	173
Qy	203	LIYTFGSAGNMMSOMITGLRYLSGINVLONCEVALSYKKKVADYIDATTEKSEGVPE	262
Db	174	--YT--KPGENFTSLIR-----AANHDKIKEILDADYLLDTPVP--	210
Qy	263	KVRLTERPENLSSNSEILDWDIYOYKFLAERGDVQIQVSLGQLHLGRKGI--DQDYK	320
Db	211	-----CSEKKKVAIYQE--LANKQDPAELKLG--FMNEHGLLPPRDYHK	251
Qy	321	ALHYFLKAAKAGSANAMAFIGKMYLEGNAAYPONNATAFKYFSMAASKGNALGHLGLL	380
Db	252	AEWYQKSAEOGNPIAQVLLGNMYTYLGR--GVDRDVNKAIDMLKKSAQNYVPKVGGLGI	310
Qy	381	YFHCKGVPNLNVAEALKYFOKAEEKGWPDAQOLQGFMYYSGSGINWDYKLAIFYFLASQS	440
Db	311	YEMSKH---NYPEAKKMYTLASKFHFNPOALYNLGLMYEYKGVKSDPOKAFRLYKDAAQN	367
Qy	441	GQPIAIIYVLAKMYATGTVGVVRSCTFAVELYKGVCELGHLHWAKEKFIATAYFAYKDGLDLSLV	500



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:35:13 ; Search time 1542.33 Seconds  
(without alignments)  
22141.281 Million cell updates/sec

Title: US-09-714-882-1  
Perfect score: 2070  
Sequence: 1 atgaagccctgtctctgtt.....ttagaataccatgggtag 2070

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2069.6	100.0	2070	6	AX148133	AX148133 Sequence
2	2053.6	99.2	2067	6	AX148135	AX148135 Sequence
3	1650.2	79.7	1773	6	AX148137	AX148137 Sequence
4	1593.8	77.0	2129	9	HSN802045	AX137678 Homo sapi
5	1403.6	67.8	1500	6	AX148141	AX148141 Sequence
6	1403.6	67.8	1728	6	AX148145	AX148145 Sequence
7	1403.6	67.8	1731	6	AX148143	AX148143 Sequence
8	1254.6	60.6	1257	6	AX148139	AX148139 Sequence
9	631.4	30.5	2382	6	E26632	E26632 TSA305 gene
10	631.4	30.5	3473	9	AF052059	AF052059 Homo sapi
11	631.4	30.5	7885	6	E26633	E26633 TSA305 gene
12	612.2	29.6	2441	10	AB048195	AB048195 Mesocric
13	612.2	29.6	2441	10	AB048195	AB048195 Mesocric
14	607.2	29.3	3663	10	AF063095	AF063095 Mus muscu
15	510.4	24.7	557	6	AX079694	AX079694 Sequence
16	444	21.4	2144	6	AX136309	AX136309 Sequence
17	368.4	17.8	4264	3	AF181652	AF181652 Drosophill
18	331.6	16.0	2406	3	CEU50829	U50829 Caenorhabdi
19	179.6	8.7	100272	9	HSJ842G6	AL109657 Human DNA
20	172.4	8.3	156813	9	HSJ631M13	AL117333 Human DNA
21	163.6	7.9	37434	3	CEP45B3	Z78063 Caenorhabdi
22	163.2	7.9	3613	3	CEU50828	U50828 Caenorhabdi
23	148	7.1	2322	8	AY039982	AY039982 Arabidops
24	143.2	6.9	453	10	AF304855	AF304855 Rattus no
25	117	5.7	1672	6	AX092286	AX092286 Sequence
26	115.4	5.6	109151	2	AC017637	AC017637 Drosophill
27	115.4	5.6	161278	3	AC008223	AC008223 Drosophill
28	115.4	5.6	239327	3	AE003744	AE003744 Drosophill
29	112.2	5.4	190153	3	AC008201	AC008201 Drosophill
30	90.2	4.4	571	6	AR107131	AR107131 Sequence
31	90.2	4.4	571	9	HSU11037	U11037 Homo sapien
32	85.4	4.1	482	9	AB0247AS15	AB024761 Homo sapi
33	85.4	4.1	167339	2	CNS01DSD	AL121769 Homo sapi
34	85.4	4.1	174788	9	CNS01DWD	AL136039 Human chr
35	85.4	4.1	180254	2	AC084001	AC084001 Homo sapi
36	85.4	4.1	182799	2	AC073526	AC073526 Homo sapi
37	83.2	4.0	34001	6	AX067443	AX067443 Sequence
38	82.4	4.0	761	11	CNS06KKG	AL402998 T7 end of
39	81.4	3.9	70511	2	AC091083	AC091083 Homo sapi
40	80.8	3.9	231972	2	AC068055	AC068055 Homo sapi
41	80.6	3.9	156550	2	AC015830	AC015830 Homo sapi
42	80.2	3.9	6644	6	E23356	E23356 Virus vecto
43	80.2	3.9	7218	6	I66494	I66494 Sequence 14
44	80.2	3.9	7372	6	E23357	E23357 Virus vecto
45	80.2	3.9	7797	6	E23355	E23355 Virus vecto

ALIGNMENTS

RESULT 1  
AX148133  
LOCUS AX148133  
DEFINITION Sequence 1 from Patent WO0136636.  
ACCESSION AX148133  
VERSION AX148133.1  
KEYWORDS GI:14347059  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2070)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 1 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES  
source  
1..2070  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 701 a 379 c 452 g 537 t 1 others  
ORIGIN

Query Match 100.0%; Score 2069.6; DB 6; Length 2070;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2070; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagcccttgctctgttaataagagatatgtataatctcttggtggtcacaataaaact 60  
DB 1 ATGAAGCCCTTGCTCTGTGTAATAGAGATATTGATAATCTTGGGGTCACAATTAAGAACT 60  
QY 61 atcaagcagagaaacataataaaagacaaaagaaagaaatgtccaccacaggtatca 120  
DB 61 ATCAAGCAGAGAAACATATATAAGACAAAAGAAAGAAATGTCACCACACAGGTATCA 120  
QY 121 gtgaagaaatcaaaacataattttatcacacatatgtggaacaaagaacatctagtaatgta 180  
DB 121 GTGAAGCAATCAACAATATTTATCACACATATTGGAACAAAGAACATCTAGTAATGTA 180  
QY 181 atcaataaaagagaaatctcctggagaaaagaaagaatacaacgtaaaaataagaataaaa 240  
DB 181 ATCAATAAAGAGAAATCTCTGGAGAAAAGAAAGAAATCAACGTAAAAATAAGAAATAAA 240  
QY 241 ggaattcaaaataaagatatcttgaaagagaaataaagaatcatttacaagaacagcagag 300  
DB 241 GGAATTCMAAATAAAGATATCTGAGAGANAATAAGAAATCAATTTACAAAAGCAGAG 300  
QY 301 aaaaaatttacagatgaagagaccagctatttaagatgggcatcaaggtttctccagcag 360  
DB 301 AAAAAATTTACAGATGAAGGAGACCAGCTATTTAAGATGGGCATCAAGGTTCTCCAGCAG 360  
QY 361 tctaaagccaaaacaaaagaaagagcctacctctcttttgcacaaagcagctgacatg 420  
DB 361 TCTAAAGCCAAAACAAAAGAAAGAGCCCTACCTACTTTTGGCCAAAGCAGCTGACATG 420  
QY 421 ggaacttgaaagctatgagaaaaatggctgcgcgtttgtctatttggaaatttggcgtg 480  
DB 421 GGAACCTTGAAAGCTATGAGAAAAATGGCTGACGCTTTGCTATTGTGAAATTTGGCGTG 480  
QY 481 caaaatataacagcagctatccaattatatagtccttggcttaagaaagagatcatgaaa 540  
DB 481 CAAAATATAACAGCAGCTATCCAATTTATATGATGCTCTGGCTTAAAGAAAGGATCATGTAAA 540  
QY 541 gcccaaaacgcataggaattttgtctctcttctgaataggaatggaatgatgatcaagct 600  
DB 541 GCCCAAAACGCTTAGGATTTTGCTCTCTATGGAATAGGAATGGAATATGATCAAGCT 600  
QY 601 agggcactgatatttacaccttggaaagtgtcggaggaacacatgatgtcccagatgatt 660  
DB 601 AAGGCACCTGATATATTACACCTTTGGAAGTGTGGAGGAAACATGATGTCCCAGATGATT 660  
QY 661 ttgggttacagatatttgcgggaatcaatgttctacagaattgtgaagtgtccctaaagt 720  
DB 661 TTGGGTACAGATATTGTGCGGAATCAATGTCTACAGAAATGTGAAGTTGCCCTAAAGT 720  
QY 721 tattaagaagaagtcagcatattatgtgcacacatttgaataaagtgaaagtggttcca 780  
DB 721 TATTACAAGAAAGTGCAGATATATTGTGTGACACATTTGAAAAAAGTGAAGGTGTCCA 780  
QY 781 gtgaaaaaagtgagactaacgaaagacctgaaaaatctgagttctaaacagtgagatttg 840  
DB 781 GTGAAAAAAGTGAGACTAACGGAAAGACCTGAAAAATCTGAGTTCTAACAGTGAGATTTTG 840  
QY 841 gattggacatatcaaatactataaatttttggcagaagaagggagatgttcagatacaa 900  
DB 841 GATTGGGACATATACCAATACTATAAATTTTGGCAGAAAGAGGAGATGTTCCAGATACAA 900  
QY 901 gtctctcttggaacaattacatctaatttggcaggaaggtctctagatcaggattactacaaa 960  
DB 901 GTCTCTCTTGGACAATTACATCTAATTGGCAGAAAGGCTTAGATCAGGATTAACAAA 960  
QY 961 gcattacactcttcttaaaggcagcaaggccggagtgcaaatgcccatttata 1020

DB 961 GCATTACACTACTTCTTAAGGCAGCAAGGCGGAGGTGCAAAATGCCATGGCATTTATA 1020  
QY 1021 ggaagatgtatttagaggggaatgctgcgtgcccacaaaataacgtactactgcttcaag 1080  
DB 1021 GGAAGATGTATTATTAGAGGGAATGCTGCGGTGCCGCAAAAATAACGTACTACTGCTTCAAG 1080  
QY 1081 tacttttccatggcagcagcaaggaaggaatgaatcgaatcggccttcacatgggcttggtctctt 1140  
DB 1081 TACTTTTCCATGGCAGCCAGTAAGGGCAATGCAATCGGCTTCAATGGGCTTGGCTCTCTT 1140  
QY 1141 tactttcatggaaaaagaggttcccttgaattatgcrcraagcacttaaatactttcagaaa 1200  
DB 1141 TACTTTTCATGGAAAAGAGAGTTCCTCTGAATATATGCCRAAGCACTTAAATACTTTCAGAAA 1200  
QY 1201 gctgggaaaaaggggtggcccgacgcacagttccagtttaggtcttcattactactctg 1260  
DB 1201 GCTGCGAAAAAGGGTGGCCGACGCACAGTTCCAGTTAGGCTTTCATGTACTACTCTGGC 1260  
QY 1261 tctggaatatggaagattataaaacttgccttcaaatatttttaccctggcattcagagt 1320  
DB 1261 TCTGGAATATGGAAGGATTATAAACTTGCTTCAAAATATTTTACCTGGCATCTCAGAGT 1320  
QY 1321 gggcagccctcgcatttattctctggcacaagatgtatgcaacaggaacagagtagta 1380  
DB 1321 GGGCAGCCCTCGCCATTTATTATCTGGCCCAAGATGTATGCAACAGGAACAGAGTAGTA 1380  
QY 1381 agatcatgcagaactgctggagcttataaagggtgctggaactaggccactgggct 1440  
DB 1381 AGATCATGCAGAACTGCTGTGGAGCTTTATAAAGGTGCTGTGAACTAGGCCACTGGGCT 1440  
QY 1441 gagaaatcctgcacagcttactttgccttaaggaatgggtgatatagattctctctgt 1500  
DB 1441 GAGAAATTCCTGACAGCTTACTTTTGCTTATAAGGATGGTGATATAGATTCTTCTCTGTT 1500  
QY 1501 cagtaTgcactgcttgcaaaaatggggtatgaagtagctcaagcaaatccagcattcatt 1560  
DB 1501 CAGTATGCACACTGCTTGCAAAATGGGATTAAGTAGTCTCAAGCAATTCAGCATTCATT 1560  
QY 1561 ttggaaatctaaaaaggctaaacttcttgaaaaagagaagatgtatccaatggcgtctctc 1620  
DB 1561 TTGGAATCTTAAAAAGGCTTAACATTTCTGAAAAAGAGAAGATGTATCCAATGGCGTTCTC 1620  
QY 1621 ctatggaatcgagctgcccattcaaggcaatgcatcttgcagagtaaaaaattggagattac 1680  
DB 1621 CTATGGAATCGAGCTGCCATTTCAAGGCAATGCTATTGTCTAGAGTAAAAAATTGGAGATTAC 1680  
QY 1681 cattactatggctatgggactaagaaagactatcaaacagcagccacacacactacagcatt 1740  
DB 1681 CATTACTATGGCTATGGGACTTAAGAAAAGACTATCAAAACAGCAGCCACACACTACAGCAT 1740  
QY 1741 gcagccacaataacacacaaacgcgaagccatgttcaattctggcttatgtatgaacac 1800  
DB 1741 GCAGCCAAACAAATACCAACCGCAAGCCATGTTCAATCTGGCTTATATGTATGAACAC 1800  
QY 1801 ggcttaggcatacaaaaagacattcacttggcagaagattgtacacatggctgctcaa 1860  
DB 1801 GGCTTAGGCATCACAAAAGGACATTCACITGGCCAGAGATTGTACGACATGGCTGCTCAA 1860  
QY 1861 acgagtcacagatggccacatacctgtctcttggcgtcatgaaactgaaaaactacgcat 1920  
DB 1861 ACGAGTCCAGATGCCACATACCTGTGCTCTTTGGCGTTCATGAAACTGGAACACTACGCAT 1920  
QY 1921 ttgctccgggatactcgttttttaatacagtttcaacaaagagatggaactggctgaaactg 1980  
DB 1921 TTGCTCCGGGATATCCTGTTTTTAAATCAGTTTCACAACGAGATGGAACCTGGCTGAAACTG 1980  
QY 1981 gacaacaccattggccacacactgggacttatttggattggcctcatgttctctgggctg 2040  
DB 1981 GACAACACCATTTGGACCACACTGGGACTTATTTGTGATTGGGCTCATTTGTTCTTGGGCTG 2040  
QY 2041 atttcttggcttagaanaatccaccatgggtag 2070  
|||||

Db 2041 ATTTTGTTCCTTAGAAATCACCATGGGTAG 2070

RESULT 2  
AX148135  
LOCUS AX148135 2067 bp DNA PAT 08-JUN-2001  
DEFINITION Sequence 3 from Patent WO0136636.  
ACCESSION AX148135  
VERSION AX148135.1 GI:14347060  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2067)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 3 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES  
source Location/Qualifiers  
1..2067  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 700 a 378 c 451 g 537 t 1 others  
ORIGIN

Query Match 99.2%; Score 2053.6; DB 6; Length 2067;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2067; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 atgaagccctgtctctgttaataagagatatattgataattcttgggggtcaacattaaaaact 60  
Db 1 ATGAAGCCCTTCTCTCTGTTAATAGAGATATTGATAAATCTTTGGGTCACAATTAATAACT 60

QY 61 atcaagcagagaacataataaaagacaaaagaaagaaatgtcacacacaggtatca 120  
Db 61 ATCAAGCAGAGAACATAATAAAGACAAAAGAAAGAAATGTCAACACACAGGTATCA 120

QY 121 gtgaacgaatacaacaattattcacacatatgtggaacaaagaacacattctgaatga 180  
Db 121 GTGAACGAATCAACATATTATTACACACATATTGGACAAAGAACATCTAGTAATGTA 180

QY 181 atcaataaaagaaatactctgtgagaaaaagaagaatcaacgttaataagaataaaa 240  
Db 181 ATCAATAAAAGAAATCTCTGTGAGAAAAGAAAGAAATCAACGTAAATAAGAAATAAAA 240

QY 241 ggaattcaaaataagagatatcttgagagagaataagaatcatttcacaaagaacagag 300  
Db 241 GGAATTCAAAATAAGATATCTTGAAGAGAAATAAGAAATCAATTTACAAAAGCAAGCAGAG 300

QY 301 aaaaattttacagatgaagagagaccagctatttaagatgggcatcaaggttctccagcag 360  
Db 301 AAAAATTTACAGATGAAGGACACAGCTATTATTAGATGGGCATCAAGTTCTCCAGCAG 360

QY 361 tctaaagcgaacaaacaaagaagcctacctactcttcttggcgaagcagctgacatg 420  
Db 361 TCTAAAAGCAAAACAAAGAAAGAAAGCTTACCTACTTTTGGCCAAAGCAGCTGACATG 420

QY 421 ggaacttgaagctatggagaaaatggctgcgcgtttgctatttggaaaattttggcgctg 480  
Db 421 GGAACCTTGAAGCTATGGAGAAAATGGCTGCAGCTTTGCTATTTTGGAAATTTGGCGCTG 480

QY 481 caaataaacagcagctatccaattatagctgctggctggaagaaagagatcatgtaaa 540  
Db 481 CAAATATATACAGACGCTATCCAAATATATAGTCTCTGGCTTAAGAGAAATCATGTAAA 540

QY 541 gcccaaacgcattaggattttgtctcttcttatggaatagggaatggaatatatgaagct 600  
Db 541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATCAAGCT 600

QY 601 aaggcaactgatattacacaccttggaggtgctggaggaacacatgatgtcccagatgatt 660  
Db 601 AAGGCACTGATATATATACACCTTTTGGAAAGTGTGGAGGAAACATGATGTCCCAGATGATT 660

QY 661 ttggggtacagatatattgtcgggaatcaatgttctacagaatttgaagtccctaaagt 720  
Db 661 TTGGGTTACAGATATTGTGCGGAATCAATGTTCTACAGAAATTGTGAAGTTGCCCTAACT 720

QY 721 tattacaagaagtggcagatttatattgtgcacacatttgaaaaaagtgaaaggtttcca 780  
Db 721 TATTACAAGAAGTGGCAGATTATATTGCTGACACATTTGAAATAAGTGAAGGTGTCCA 780

QY 781 gtgaaaaagtgaactaacgaaagacctgaaaactctgagttctaacagtgagattttg 840  
Db 781 GTGGAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 840

QY 841 gattgggacataacacataactataaattttggcagaagaaggagagatttcagatacaa 900  
Db 841 GATTGGGACATATACCAATACTATAAATTTTGGCAGAAAGAGAGATGTTTCAGATACAA 900

QY 901 gtctctttggacaaattacatttaattggcagaaaggctcagatcaggattactacaaa 960  
Db 901 GTCTCTCTTTGGACAATTACATCTAATTGGCAGAAAGGTTCTAGATCAGGATTACTACAA 960

QY 961 gcattacactactctttaaaggcagcaaaagcgggagtgcaaatcccatggcattata 1020  
Db 961 GCATTACACTACTCTTAAAGCAGCAAAAGCGGGAGTGCATAATGCCATGGCATTTATA 1020

QY 1021 ggaagaatgtatttagaggggaatgctgcctgcccgaacaaataacgactactgcctcaag 1080  
Db 1021 GGAAGAATGTATTAGAGGGGAATGCTGCGTGCSCCAAAATAACGCTACTGCTTCAAG 1080

QY 1081 tactttccatggcagcagtaagggaatgaatgaatggcctcgaatgggttgggttctctt 1140  
Db 1081 TACTTTTCCATGGCAGCCAGTAAGGCAATGCAATCGGCTTCATGGGCTGGTCTCTCTT 1140

QY 1141 tactttcagaaaaaggadttccctggaattatccraagcacttaatactttcagaaa 1200  
Db 1141 TACTTTTCTGGAAGAGGAGTTCCTCTGAAATATTGCCRAAGCACTTAAATACITTCAGAAA 1200

QY 1201 gctgcggaacaaagggtgcccgcagcagcagttccagttcagttcagttcagttcagtt 1260  
Db 1201 GCTGCGGAACAAAGGTTGCGCCGACGCACAGTTCCAGTTAGGCTTTCATGTACTCTGGC 1260

QY 1261 tctggaatatgaagattataaaacttgctctcaaatatttttaccctggcactctcagagt 1320  
Db 1261 TCTGGAATATGGAAGGATTATAAACTTGCTCTTCAATAATTTTACCCTGGCATCTCAGAGT 1320

QY 1321 ggagcagccctcagcatttattcttgccaaagatgtatgcaacaggaagagtagta 1380  
Db 1321 GGGCAGCCCCCTGCCATTTATTATCTGGCCAAAGATGTATGCAACAGGAACAGAGTAGTA 1380

QY 1381 agatcatgcagaactgctgtagagctttataaagggtgctgtagaactagccactgggct 1440  
Db 1381 AGATCATGCAGAACTGCTGTGGAGCTTTATAAGGTGTCTGTGAACCTAGGCCACTGGGCT 1440

QY 1441 gagaaattcctgacagcttactttgctctaagaagtggtgatagattctctcttctgtt 1500  
Db 1441 GAGAAATTCCTGACAGCTTACTTTTGCTATTAAGATGGTGTATAGATTTCTCTCTGTGT 1500

QY 1501 cagatgcaactgctgcagaaaatggggtatgaatgagtagtcaaaagaattcagcattcatt 1560  
Db 1501 CAGTATGCACTGCTGTCAGAAATGGGGTATGAAGTAGCTCAAAAGCAATTCAGCATTCATT 1560

QY 1561 ttggaattcaaaaggctcaactcttgaaaaagagaagatgtatccaatggcgctcttc 1620  
Db 1561 TTGGAATCTAAAAGGCTAACATTTCTTGAANAAGAGAAGATGTATCCATGGCGCTCTC 1620

QY 1621 ctatggaattcgagctgccattccaaggcaatgcatttctgtagagtaaaaaattggagattac 1680  
Db 1621 CTATGGAATCGAGCTGCCATTCAAAGCAATGCATTTGTCTAGAGTAAATAAATTTGAGATTAC 1680

QY 1681 cactactatgctatgggactgaagaagcattcaaaagcagcagccacacactacagcatt 1740  
Db 1681 CATTACTATGGTATGGGACTAAGAAAGACTATCAAAACAGCAGCCACACTACAGCATTT 1740

```
QY 1741 gcagccaaataaccacaaacgcgaagccatgttcaatctgcttatgtatgaacac 1800
|||||
Db 1741 GCAGCCAAATAACCAACACGCGCAAGCCATGTTCAATCTGGCTTATATGATGAACAC 1800
|||||
QY 1801 ggcctaggcatcacaagagacattccacttggcgcagaagattgtacgacatggctgctcaa 1860
|||||
Db 1801 GCCTTAGGCATCACAAAGACATTCACCTGGCCAGAAAGATTGTACGACATGGCTGCTCAA 1860
|||||
QY 1861 acgagtcagatgccacataacctgtctcttgcgcgtcatgaaactggaactacgcat 1920
|||||
Db 1861 ACAGTCCAGATGCCACATACCTGTGCTTTGGCCGTCATGAAACTGGAAACTACGCAT 1920
|||||
QY 1921 ttgctcgggatactcgtgttttataatcagttccacaacagatggaactggctgaaactg 1980
|||||
Db 1921 TTGCTCGGGATATCCTGTTTTTAAT---TTCACAACGAGATGGAACCTGGCTGAAACTG 1977
|||||
QY 1981 gacaacaccattggaccacactgggacttattgtgattggcctcattgttccctgggctg 2040
|||||
Db 1978 GACAACACCATTGGACCACACTGGGACTTATTTGTGATTGGCTCATTTGCTGGGCTG 2037
|||||
QY 2041 atttgttcttagaataccactgggtag 2070
|||||
Db 2038 ATTGTTGCTGTAGAAATCACCATGGGTAG 2067
|||||

RESULT 3
AX148137
LOCUS AX148137 1773 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 5 from Patent WO0136636.
ACCESSION AX148137
VERSION AX148137.1 GI:14347061
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1773)
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.
TITLE Human notch ligand proteins and polynucleotides encoding the same
JOURNAL Patent: WO 0136636-A 5 25-MAY-2001;
Lexicon Genetics Incorporated (US).
FEATURES
source
1..1773
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 607 a 313 c 388 g 464 t 1 others
ORIGIN
Query Match 79.7%; Score 1650.2; DB 6; Length 1773;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atgaagcccttgcctgttaagatattgataattcttctggggtcacaaattaaaact 60
|||||
Db 1 ATGAAGCCCTTCTCTGTCTTAATAGAGATATTGATAATTCTTGGGGTCACAAATTAATACT 60
|||||
QY 61 atcaagcagaggaacataataaagacaaagaaagaaagaaatgtccacacaggtatca 120
|||||
Db 61 ATCAAGCAGAGACATAATAAAGACAAAGAAAGAAAGAAATGTCCACACAGGTATCA 120
|||||
QY 121 gtgaacgaatacaacaatatttatcacacatatgttgaaacaaagaacatctagtaatga 180
|||||
Db 121 GTGAACGAATCAACAATATTTATCACACATATTGGAACAAGAATCATCTAGTAATGTA 180
|||||
QY 181 atcaataaagagaataactcctcgtgagaaaaagaagaatcaacgtaaaataagaataaaa 240
|||||
Db 181 ATCAATAAAGAGAATAACTCTCTCGAGAAAAAGAAAGAAATCAACGTTAAATAAGATAAAA 240
|||||
QY 241 ggaattcaaaataaagatatcttgaagagaataagaatacatttcaaaaagcaagcagag 300
|||||
Db 241 GGAATTCAAAATAAAGATATCTTGAAGAGAAATAAGATATCTTACAAAAGCAAGCAGAG 300
|||||
```

```
QY 301 aaaaattttacagatgaaggaacacagctatttaagatgggcatcaagttctccagcag 360
|||||
Db 301 AAAAATTTTACAGATGAAGGAGACAGCAGCTATTATAGATGGGCATCAAGGTTCTCCAGCAG 360
|||||
QY 361 tctaaaagcacaacaaagaaagcctacctactcttcttggccaaagcagctgacatg 420
|||||
Db 361 TCTAAAAGCCAAACAAACAAAGAAAGCCTACCTACTTTTTTGGCCAAACAGCAGCTGACATG 420
|||||
QY 421 ggaacttgaaagctatggagaaaaatggctgacgcgttctgtctatttggaaaattttggcgtg 480
|||||
Db 421 GGAACACTTTGAAAGCTATGGAGAAAAATGGCTGACGCTTTGCTATTATTTGGAAATTTTGGCGTG 480
|||||
QY 481 caaataataacagcagctatccaattatagctccttgctgctaaagaagatcatgtaaa 540
|||||
Db 481 CAAATATATAACAGCAGCTATCCAAATATATAGAGTCTCTGGCTTAAAGAGAGATCATGTAAA 540
|||||
QY 541 gcccaaacgcattaggatttttctctcttcttgaatagggaatggaatatgatacagct 600
|||||
Db 541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATCAAGCT 600
|||||
QY 601 aaggcactgatataatacacaccttggaaagctgctgagagaaacatgatgtcccagatgatt 660
|||||
Db 601 AAGGCACCTGATATATTACACCTTTTGGAGTGTGGAGGAAACATGATGTCCAGATGATT 660
|||||
QY 661 ttggggtacagatatttgcgggaatcaatgttctacagaattgtgaagtgcctcaagt 720
|||||
Db 661 TTGGGGTACAGATATTGTGCGGAATCAATGTTCTACAGAAATGTGAAGTTCGCCCTAAGT 720
|||||
QY 721 tattcaagaagagtgccagatttatattgctgcacacatttggaaaaagtgaaagtggttcca 780
|||||
Db 721 TATTACAAGAAAGTGGCAGATTATATTGCTGACACATTTTGAAAAAAGTGAAGGTGTCCA 780
|||||
QY 781 gtgaaaaagtgagactaacggaaagacctgaaatctgagttctaaacagtggattttg 840
|||||
Db 781 GTGAAAAAGTGAGACTAACGGAAAGACCTGAAAAATCTGAGTCTTAACAGTGAAGTTTG 840
|||||
QY 841 gattggacataacacataactataaattttggcagaagaggagatttcagatacaca 900
|||||
Db 841 GATTGGACATATACCAATACTATAAATTTTGGCAGAAAGAGAGATGTTCAGATACAA 900
|||||
QY 901 gtctcttggacaattacatctaattggcaggaaggtcctagatcagagattactacaaa 960
|||||
Db 901 GTCTCTTGGACAATTACATTAATTGGCAGAAAGGTCTAGATCAGGATTACTACAAA 960
|||||
QY 961 gcttaccactactcttaaggcagcaaaagccggagtgcaaatgccatgcatattata 1020
|||||
Db 961 GCATTACACTACTTCTTAAGGCACAAAGCCGGAGTGCANATGCCATGCAATTTATA 1020
|||||
QY 1021 ggaagatgtatttagaggggaatgctgccgtgccgcaaaaataacgctactgccttcaag 1080
|||||
Db 1021 GGAAGATGTATTTAGAGGGGAATGCTGCCGTGCCGCAAAAATAACGCTACTGCTTCAAG 1080
|||||
QY 1081 tactttccatggcagccagtaaggcgaatgcaatgcgccttcacgttgccttggttcttt 1140
|||||
Db 1081 TACTTTTCATGGCAGCCAGTAAGGCAATGCAATCGGCTTTCATGGGCTTGGCTCTCT 1140
|||||
QY 1141 tactttcatggaagagaggttcccctgaattatgccraagcacttaatactttcagaaa 1200
|||||
Db 1141 TACTTTTCATGGAAGAGAGTTCCTCCTGAATATATGCCRAGCACTTAATATCTTTCAGAAA 1200
|||||
QY 1201 gctgcgaaaaaggggtggccgcagcagttcccagttcaggttaggtctactactctggc 1260
|||||
Db 1201 GCTGCGAAAAAGGGTGGCCCGCAGCACAGTTCCAGTTCCAGTTTCACTGACTCTGSC 1260
|||||
QY 1261 tetggaatggaaaggattataaaacttgccttcaaatatttttaacctggcatctcagagt 1320
|||||
Db 1261 TCTGGAATATGGAAGGATTATAAACTTGCCTTCAAAATATTTTACCTGGCATCTCAGAGT 1320
|||||
QY 1321 gggcagccctcctgcatttattctgtgcgaagatgtatgcaacaggaacagagtagta 1380
|||||
Db 1321 GGGCAGCCCTCGCCATTATTATCTGCGCAAGATGTATGCAACAGGAAACAGAGTAGTA 1380
|||||
```



```
QY 1381 agatcatgcagaaactgctgtgagatttataaaaggtgtctgtgaaactaggccactgggct 1440
Db 1381 AGATCATGCAGAACTGCTGTGAGGTTTATAAAGGTGTCTGTGAACCTAGGCCACTGGGCT 1440
QY 1441 gagaaattcctgcagacttaacttgcctataaaggatggtgatagattcttctctgtt 1500
Db 1441 GAGAAATTCCTGCAGACTTACTTTGCCCTATAAGGATGATGATATAGATTCTTCTCTGTT 1500
QY 1501 cagtgactctgtgcagaaatggggtatgaagttagctcaagcaattcagcattcatt 1560
Db 1501 CAGTATGCACCTCTTGCGAGAAATGGGTATGAAGTAGCTCAAGCAAAATTCAGCATTCATT 1560
QY 1561 ttggaaatctaaagaagctaaacttcttgaagaagaagatgatcccaatgctgtcttc 1620
Db 1561 TTGGAATCTAAAGAGCTAAACATTTTGAAGAGAGAGATGATATCCAAATGCGCTTCTC 1620
QY 1621 ctatggaatcgactgctgcaattcaaggcaatgcatttg 1657
Db 1621 CTATGGAATCGAGCTGCCATTCAAGGACATTCACTTG 1657

RESULT 4
LOCUS HSM802045 2129 bp mRNA PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp434C1826 (from clone DKFZp434C1826).
ACCESSION AL137678
VERSION AL137678.1 GI:6807900
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2129)
AUTHORS Ansorge,W., Wirkner,U., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp434C1826) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
source
1..2129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434C1826"
/clone.lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="testis"
polyA_signal 2100..2105
polyA_site 2118
BASE COUNT 713 a 413 c 486 g 517 t
ORIGIN

Query Match 77.0%; Score 1593.8; DB 9; Length 2129;
Best Local Similarity 91.2%; Pred. No. 5.8e-306;
Matches 1839; Conservative 1; Mismatches 3; Indels 174; Gaps 7;

QY 57 aactatcaagcagagggaacataataaaagacaaaaggaagaaatgtcacacacagct 116
Db 194 AGCTATCAAGCAGAGGAACATAATAAAAGACAAAAGGAAGAAATGTCCACCACAGGT 253
QY 117 atcagtgaaacgaatacaacaattattcacacatttgggaacaaagacatctagtaa 176
Db 254 ATCAGTGAACGAATCAACAATATTTATCACACATATTGGGAACAAAGAACATCTAGTAA 313
```

```
QY 177 tgtaatcaataaaagagaaaaatctctctggagaaaaagaaatcaacgtaaaaatgaagaat 236
Db 314 TGTAAATCAATAAAGAGAAAAATCTCTCTGGAGAAAAAGAAATCAATGTAAAAATAAGAAAT 373
QY 237 aaaaagaattcaaaaataaagatactttgaagagaaataaagaatcatttcaaaaagaacac 296
Db 374 AAAAGGAATTCAAAATAAAGATATCTTGAAGAGAAATAAGAATCATTTTACAAAGCAAGC 433
QY 297 agagaaaaattttacagatgaaggagaccagctattttaagatgggcatacaaggttctcca 356
Db 434 AGAGAAAAATTTTACAGATGAAGGAGACCAGCTATTTAAGATGGGCATCAAGGTTCTCCA 493
QY 357 gcagtctaaaagcccaaaaaacaaaaaagaagaagcctcactcttbtgccaagaagcagctga 416
Db 494 GCAGTCTAAAGGCCAAAAACAAAAAGAGAGAGCTACCTACTTTTGGCCAAAGCAGCTGA 553
QY 417 catggaaaacttgaagctatggaaaaatggctacgcttcttcttcttcttcttcttcttctg 476
Db 554 CATGGAAAACCTTGAAGCTATGGAAAAATGGCTGACGCTTTGCTATTTTGGAAAAATTTGG 613
QY 477 cgtgcaaaataaacagcagctatccaattatgatgctcttggctggaagaagat-cat 535
Db 614 CGTGCAAAATATAACAGCAGCTATCCAATATATGAGTCTCTGGCTTAAGAGAGGATCCAT 673
QY 536 gtaaaagcccaaacgcatttaggatttttctctcttcttcttcttcttcttcttcttcttcttct 595
Db 674 GTAAAGCCCCAAAACGCTATGAGGATTTTCTCTCTTATGAGTAGGAAATGGAATATGATC 733
QY 596 aagctaaggcactgatataattacacaccttgggaagtgctggaagaacacatgatgccacaga 555
Db 734 AAGCTAAGCAGCTGATATATTACACTTTTGGAAAGTCTGGAGGAAACATGATGTCCACAGA 793
QY 656 tgatttgggggtacagatatgttgcgggaatcaatgttctacagaatgtgaagtgtgcc 715
Db 794 TGATTTTGGGGTACAGATATTTTGTGGGAATCAATGTTCTACAGAAATTTGAAAGTTGCC 853
QY 716 taagtattacaagaagtggcagatttatattgctgacacatttggaaaaagtgaaaggtg 775
Db 854 TAAGTTATTACAAGAAAGTGGCAGATTATATTGCTGACACATTTTGAAAAAGTGAAGGTG 913
QY 776 ttccagtggaagaaagtgaagtaacggaagacccctgaaactgtgagttcttaacagtgaga 835
Db 914 TTCCAGTGGAAAAAGTGAGACTAAACGAAAGACCTTGAAAAATCTGAGTTCTTAACAGTGAGA 973
QY 836 ttttgattggacatataaccataactataaaattttggcagaagaagagagatgttccaga 895
Db 974 TTTTGGATTGGACATATACCAATACTATAAATTTTGGCAGAAAGAGAGATGTTTCA 1033
QY 896 tacaagtc-tctcttggacaattacatctaatgtggcaggaaggtctagatcaggattac 954
Db 1034 TACAAGTCTTCTCTTGGACAATTACATCTAATTTGGCAGGAAAGGTCTAGATCAGGATTAC 1093
QY 955 tacaagaattacactactacttcttaaaaggcagcaaaagccgggagtgcaaatgccatggca 1014
Db 1094 TACAAGCATTTACCTACTTCTTTAAAGGCGCAAGAGCCGGGAGTGCAAAATGCCATGGCA 1153
QY 1015 ttatataggaagatgtatttagaggggaatgctgcggtgcccgaataaacactactgcc 1074
Db 1154 TTTATAGGAAGATGTATTTTAGAGGGGAATGCTGCGGTGCCGCAAAATTAACGCTACTGCC 1213
QY 1075 ttcaagtacttttccatggcagccagtaagggaatgaatgcattcgttcattcgttcgttgg 1134
Db 1214 TTCAAGTACTTTTCCATGGCAGCAGTAAGGGCAATGCAATCGGCCTTCATGGGCTTGGT 1273
QY 1135 cttcttacttctcatggaaaaagaggttccctgaattatgcacaaagcacttaaaacttt 1194
Db 1274 CTCTTTTACTTTTATGGAAAAAGAGTTCCTCCCTGAATTTATGCCGAAGCACTTAATACTTT 1333
QY 1195 cagaagctcgcgaaaaaggggtggcccgacgcacagttccagttcagttaggtcttactgtactac 1254
Db 1334 CAGAAAGCTCGCGAAAAAGGGT-GCCCGACGCACA-TTCCAGTTAGGCTTCATGTACTAC 1391
```

QY	1255	tctgctctggaatggaagattataaactgcttcaaatatttttacctggcatct	1314
Db	1392	TCTGGCTCTGGAATATGGAAGATTATAAACTTGCCTTCAAAATATTTTACCTGGCATCT	1451
QY	1315	caagatgggcagccctcgccatttattatctcggcaagatgtatgcacaagacagga	1374
Db	1452	CAGAGTGGCGAGCCCTCGCCATTTATTATCTGGCCAAAGATGTATGCAACAGGAACAGGA	1511
QY	1375	gtagtaagatcatgcagaactcgtggaagctttataaagggtctgtgaaactaggccac	1434
Db	1512	GTAGTAAGATCATGCAGAACTGCTGTGGAG-----	1541
QY	1435	tgggctgagaaattcctgacagcttacttctgctataaggatgggtgatagattctct	1494
Db	1542	-----	1541
QY	1495	cttgctcagtatgcactgctgcagaaaatggggatgaagtagctcaaaagcaattcagca	1554
Db	1542	-----	1541
QY	1555	ttcattttggaattctaaaaggcttaacattcttgaagaagagaagatgtatccaatggcg	1614
Db	1542	-----AAAGGCTAACATCTTGAAGAGAGAAAGATGTATCCAATGGCG	1585
QY	1615	cttctcctatggaatcgagctgcccattcgaagcaatgcatttgcctagatgaataattgga	1674
Db	1586	CTTCTCCTATGGAATCGAGCTGCCATTCAAGGCAATGCAATTTGCTAGAGTAAATAATTGA	1645
QY	1675	gattaccattactatgctatgggactgaagaagactatcaaacagcagccacacactac	1734
Db	1646	GATTACCATTACTATGGCTATGGGACTAAGAAGAGACTATCAAACAGCAGCCACACTAC	1705
QY	1735	agcattgcagccacaataaccacaacgcgcgaagccatttccaatcgttcttatgtat	1794
Db	1706	AGCATTGCAGCCACAATAATACCACAACGCGCAAGCCATCTTCAATCTGCTTATATGAT	1765
QY	1795	gaacaggttaggcatacaagaagacattcactctggccaggaagattgtacgacatggct	1854
Db	1766	GAACAGGGTTAGGCATACAAAGGACATTTCACTTGGCCAGAGATTTGACGACATGGCT	1825
QY	1855	gctcaaacagctccagatccacataccctgctcttgccttcgctatgaactggaact	1914
Db	1826	GCTCAAACAGTCCAGATGCCACATACCTGTCTTTTGGCGTATGAACCTGGAACCT	1885
QY	1915	acgcatttgcctcgggatactcctgtttttaaactcagttccacaacagatggaaactggctg	1974
Db	1886	ACGCATTTGCTCGGGATATCTGTTTTTAAT---TTCAACAACGAGATGGAACCTGGCTG	1942
QY	1975	aaactggacaacacacattggacacactgggacttatttgccttgcctcattgttct	2034
Db	1943	AAACTGGACAACACCATTTGGACACACTGGGACTTATTGTGATTGGCCTCATTTGCTCT	2002
QY	2035	gggctgatttggct-gcttagaataaccatgggtag	2070
Db	2003	GGGCTGATTTGTGGCTTAGAAATCACCATGGGTAG	2039
RESULT	5		
AX148141			
LOCUS	AX148141	1500 bp	DNA
DEFINITION	Sequence 9 from Patent WO0136636.		PAT
ACCESSION	AX148141		
VERSION	AX148141.1	GI:14347063	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 1500)		
AUTHORS	Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.		
TITLE	Human notch ligand proteins and polynucleotides encoding the same		
JOURNAL	Patent: WO 0136636-A 9 25-MAY-2001; Lexicon Genetics Incorporated (US)		

FEATURES	Location/Qualifiers	
source	1..1500	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
BASE COUNT	537 a 254 c 328 g 380 t	1 others
ORIGIN		
Query Match	67.8%; Score 1403.6; DB 6; Length 1500;	
Best Local Similarity	100.0%; Pred. No. 2.8e-268;	
Matches 1404; Conservative	0; Mismatches	0; Indels
0; Gaps	0;	
QY	1 atgaagcccttctctgttaataagatattgataattcttgggggtcacaattaaaaact	60
Db	1 ATGAAGCCCTTCTCTCTGTTAATAGAGATATTGATAATTCTTTGGGGTCACAATTAATAACT	60
QY	61 atcaagcagcagaagacataataaaagacaaaaggaagaaaatgtcaccacacaggtatca	120
Db	61 ATCAAGCAGAGGAACATAATAAAGACAAAAGGAAGAAATGTCCACCACAGGTTATCA	120
QY	121 gtgaacgaaaatacaacaattatttaccacatatgttggaaacaaagacatcttagtaatga	180
Db	121 GTGAACGAATCAACAATATTATTATCACACATATTGGAAACAAGAACAATCTAGTAATGTA	180
QY	181 atcaataaagagaataatctcctggagaaaagagaatacaacgtaaaaataagaataaaa	240
Db	181 ATCAATAAAGAGAAAATCTCCTGGAGAAAAGAGAATCAACGTAATAATAAGATAAAA	240
QY	241 ggaattcaaaaataagatatcttgaagagaaaataagaatacatttcaaaaagcagcagag	300
Db	241 GGAATTCAAAATAAAGATATCTTGAAGAGAAAATAGAAATCATTTACAAAAGCAAGCAGAG	300
QY	301 aaaaattttcacagatgaagagaccagctatttaagatgggcatcaagggtctctccagcag	360
Db	301 AAAAATTTTACAGATGAAGAGACCAGCTATTTAAGATGGGCATCAAGGTTCTCCAGCAG	360
QY	361 tctaaaagcgaacaaaagaaagcctacctacttttgcgaagcagctgacatg	420
Db	361 TCTAAAAGCCAAAACAAAAGAAAGAGCGCTACCTACTTTTGGCCAAAGCAGCTGACATG	420
QY	421 gaaacttgaagctatggagaaaatgctgacgtttgctatttggaaatttggcgtg	480
Db	421 GGAACCTTGAAGAGCTATGGAGAAAATGGCTGACGCTTTCCTATTGGAAATTTTGGCGTG	480
QY	481 caaaataaacgcagctatccaattatatagctccttgcttaagaaagcagctgacatg	540
Db	481 CAAATAATAACAGCAGCTATCCAAATTATATAGTCTCTGGCTAAAGAAGCATCATGTAAA	540
QY	541 gcccaaacgcattaggattttgtcttcttatggaatagggaatggaatgatcatcaagct	600
Db	541 GCCCAAAACGCATTAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATCAAGCT	600
QY	601 aaggcactgatattacacctttgggaagtgctggaggaacacatgatgtcccagatgatt	660
Db	601 AAGGCACGTATATATACACCTTTGSAAGTGTGGAGGAACATGATGTGCCAGATGATT	660
QY	661 ttgggttacagatatttgcgggaatcaatgttctacagaatttggaagtgccttaagt	720
Db	661 TTGGGGTACAGATATTTTGGGGGAATCAATGTTCTACAGAAATTTGGAAGTTGCCCTAAGT	720
QY	721 tattacaagaagtgagcagattatttgcctgcacacatttgaaaaagtgaaggtgtcca	780
Db	721 TATTACAAGAAAGTGGCAGATTATATTGCTGCACACATTTGAAAAGATCAAGGTGTGTCCA	780
QY	781 gtcgaaaaagtgaagcgaacgaagccctgaaatcctgagttcttaacagtgagatttgc	840
Db	781 GTGGAAGAAAGTGAGACTTAACGGAAAGACCTGAAAATCTGAGTTCTAACAGTGAAGTTTG	840
QY	841 gattggacatacaataactactataaattttggcagaagaagagagatttcagatcaca	900
Db	841 GATTGGACATATACCAATACTATAAATTTTGGCAGAAAGAGGAGATGTTTCAGATACAA	900
QY	901 gtctctcttggacaattacatcaatattggcaggaaggtctagatcaggtactactacaa	960

```
Db 901 GTCTCTCTGGACAAATTACATCTAATTGGCAGGAAGGCTCTAGATCAGGATTTACTACAAA 960
Qy 961 gcattacactactctttaaagcagcaagcccgaggatgcaaatgccatgcatttata 1020
Db 961 GCATTACACTACTTCTTAAAGGCAGCAAAAGCCGGGAGTGCAATGCCATTTATA 1020
Qy 1021 ggaagatgtatttagaggggaatgctccgtgccgcaaaaataacgctactgccttcaag 1080
Db 1021 GGAAGATGTATTTAGAGGGAAATGCTGCCGTCCGCAAAATAACGCTACTGCCCTTCAAG 1080
Qy 1081 tacttttccatggcagcagtaagggcaatgcaatgcgccttcattggcttctt 1140
Db 1081 TACTTTTCCATGGCAGCCAGTAAGGGCAATGCAATCGGCCCTTCATGGGCTTGGTCTTCTT 1140
Qy 1141 tacttctatgaaagaggttccctgaattatgcccraagcacttaatactttcacgaaa 1200
Db 1141 TACTTTCATGGAAAGAGTTCCTCCCTGAATATGCCRAGGCACITTAATATCTTCAGAAA 1200
Qy 1201 gctcggaagaaagggtggcccgagcagctccagttcaggttccatctctggc 1260
Db 1201 GCTCGGGAAGAGGGTGGCCGAGCACAGTTCAGATTAGGCTTCATGCTACTCTGGC 1260
Qy 1261 tctggaatatggaagattataaacttgccttcaaatatttttaacctggcattcagagt 1320
Db 1261 TCTGGAATATGGAAGGATTTAAACTTGCCTTCAAAATATTTTACCTGGCATCTCAGAGT 1320
Qy 1321 gggcagccctcgccatttatctgcccgaagatgtgcaacaggaacagggagttagta 1380
Db 1321 GGGCAGCCCTCGCCATTTATTAATCTGCCAAGATGTATGCAACAGGAAGGAGTAGTA 1380
Qy 1381 agateatgcagaactgctgtggag 1404
Db 1381 AGATCATGCAGAACTGCTGTGGAG 1404

RESULT 6
LOCUS AX148145 1728 bp DNA PAT 08-JUN-2001
DEFINITION Sequence 13 from Patent WO0136636.
ACCESSION AX148145
VERSION AX148145.1 GI:14347065
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.
TITLE Human notch ligand proteins and polynucleotides encoding the same
JOURNAL Patent: WO 0136636-A 13 25-MAY-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1. 1728
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 592 a 309 c 380 g 446 t 1 others
ORIGIN
Query Match 67.8%; Score 1403.6; DB 6; Length 1728;
Best Local Similarity 100.0%; Pred. No. 2.8e-268;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgaagcccttgcctgtcttaagagattgataaattcttgggggtcacattaaaact 60
Db 1 ATGAAGCCCTTGTCTCTGTTAATAGAGATATTGATAATCTTGGGGTCCACAAATPAAACT 60
Qy 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtccaccacaggtatca 120
Db 61 ATCAAAGCAGAGGAACATAATAAAGACAAAGGAAGAAATGTCCACCACACAGGTATCA 120
Qy 121 gtgaacgaatacaacaattatttaccacatatgttggaacaaagaacatctagttagta 180
```

```
Db 121 GTGAACGAATCAACAATATTTATCACACATATTTGGAACAAAGAACATCTAGTAATGTA 180
Qy 181 atcaataaagagaaaaatctctctgagaaaaagaagaatcaacgtataaataagaataaaa 240
Db 181 ATCAATAAAGAGAAAATCTCTCTGGAGAAAAGAAGAAATCAACGTAAATATAGAATAAAA 240
Qy 241 ggaattcaaaaataagatatcttgaagagaaaataagaatacatcttcaaaaagaacagag 300
Db 241 GGAATTCAAAATAAGATATCTTTGAAGAGAAATAAGAATCATTTTACAAAAGCAACAGAG 300
Qy 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaagggttctccagcag 360
Db 301 AAAAATTTTACAGATGAAGGAGACAGCTATTTAAGATGGGCATCAAGGTTCTTCAGCAG 360
Qy 361 tctaaaagccaaaacaaaagaagaaacccctaccttacttcttgcacaaagcagctgacatg 420
Db 361 TCTAANAGCCAAACAAACAAAGAGAACCCCTACCTACTCTTTTGGCCAAAGACGTCACATG 420
Qy 421 ggaacttgaagctatggagaaaatggctgacgcttctgtctatttggaaaattttggcgtg 480
Db 421 GAAACTTTGAAAGCTATGGAGAAAATGGCTGACGCTTTGCTATTTTGGAAATTTTGGCGTG 480
Qy 481 caaaatataacagcagctatccaaattatagtgcttctggctaaagaaggatcatgtaaa 540
Db 481 CAAATATAACAGCAGCTATCCAAATATATGAGTCCCTTGGCTAAAGAAAGGATCATGTAAA 540
Qy 541 gccaaaacgcataggtatttcttctctatggaataggaatggaatggaatgataatgatac 600
Db 541 GCCCAAAACGATAGGATTTTGTCTTCTTATGGAATAGGAATGGAATATGATGATCAAGCT 600
Qy 601 aagcactgatataattacaccttggaaagtgctggaggaacacatgatgtcccagatgatt 660
Db 601 AAGCACTGATATATTACACCTTTTGGAAAGTCTGGAGGAAACATGATGTCCCAGATGATT 660
Qy 661 ttgggtcacagatatgttcgggaataatgttctacagaattgtgaagttgcccataagt 720
Db 661 TTGGGTTACAGATATTTCTCGGGAATCAATGTTCTACAGAAATTTGTAAGTTGCCCTAAGT 720
Qy 721 tattacaagaagtggcagatttatattgctgacacatttgaaaaaagtgaaggtgttcca 780
Db 721 TATTACAGAAGTGGCAGATTATATTGCTGNACACATTTGAAAAAAGTGAAAGTGTTCCTCA 780
Qy 781 gtgaaaaagtggagactaacggaagacctgaaactctgagttctaaacagtgagatttg 840
Db 781 GTGGAAGAGTGAGACTTAACGGAAAGACCTGAAATCTGAGTCTTAACAGTGAGATTTTG 840
Qy 841 gatgggcataataccaatactataataatttttggcagaaagaggagatttcagatacaa 900
Db 841 GATTGGGACATATACCAATACTATATAATTTTGGCAGAAAGAGAGAGATGTTCCAGATACAA 900
Qy 901 gtcctcttggacaattacatctaattggcaggaaggtctctagatcagattactacaaa 960
Db 901 GTCTCTCTTGGCAATTTACATCTAATTGGCAGAAAGGCTCTAGATCAGGATTTACTACAAA 960
Qy 961 gcattacactactcttaagagcagcaagccggggagtgcgcaaatgcccatttata 1020
Db 961 GCATTACACTACTTCTTAAAGGCAGCAAAAGCCGGGAGTGCAATGCCATTTATA 1020
Qy 1021 ggaagatgtatttagaggggaatgctgccgtgcccgaataataacgctactgccttcaag 1080
Db 1021 GGAAGATGTATTTAGAGGGGAATGCTGCCGTGCCGCAAAATAACGCTACTGCCCTTCAAG 1080
Qy 1081 tacttttccatggcagccagtaagggcaatgcaatcgcccttcactgggttcttctt 1140
Db 1081 TACTTTTCCATGGCAGCCAGTAAGGGCAATGCAATCGGCCCTTCATGGGCTTGGTCTTCTT 1140
Qy 1141 tactttcatgaaagaggttccctgaattatgcccraagcacttaatactttcacgaaa 1200
Db 1141 TACTTTTCAATGAAAAGGAGTTCCCTGTAATATGCCRAGGCACITTAATATCTTCAGAAA 1200
Qy 1201 gctcggaagaaagggtggcccgagcagctccagttcaggttccatctctggc 1260
```

Db 1201 GCTCGGAAAAAGGTTGGCCCGACGACAGTTCCAGTTAGGCTTCATGTACTACTCTGGC 1260  
Qy 1261 tctggaatgaagattataaaacttgccttcaaatattttaccctggcatctcagagt 1320  
Db 1261 TCTGGAATATGAAGGATTATAACTTGCCTTCAATATTTTACTGGCATCTCAGAGT 1320  
Qy 1321 gggcagccctcgccatttattcttgcccaagatgtatgcacagggaacagagtagta 1380  
Db 1321 GGGCAGCCCTCGCCATTATTATCTGGCCAAGATGTATGCAACAGGAAACAGGAGTAGTA 1380  
Qy 1381 agatcatgcagaactgctgtggag 1404  
Db 1381 AGATCATGCAGAACTGCTGTGGAG 1404

RESULT 7  
AX148143  
LOCUS AX148143 1731 bp DNA PAT 08-JUN-2001  
DEFINITION Sequence 11 from Patent WO0136636.  
ACCESSION AX148143  
VERSION AX148143.1 GI:14347064  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1731)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 11 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES  
source 1. .1731  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 593 a 310 c 381 g 446 t 1 others  
ORIGIN

Query Match 67.8%; Score 1403.6; DB 6; Length 1731;  
Best Local Similarity 100.08; Pred. No. 2.8e-268;  
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaagcccttgcctgttaagatattgataattcttgggtgcacaaataaaact 60  
Db 1 ATGAAGCCCTTGCTCTGTTAATAGAGATATTGATAATTCTTGGGTGCACAAATTTAAACT 60  
Qy 61 atcaagcagagaaacataataaagacaaaaggaagaaatgtcacacacaggtatca 120  
Db 61 ATCAAGCAGAGGAACATAATAAAGACAAAAGAAAGAAATGTCAACACAGGTATCA 120  
Qy 121 gtgaacgaatacaacaattattacacatatgttggaacaaagaacatctagttaatga 180  
Db 121 GTGACGAATCAACAAATATTATACACATATTGGACAAAGAACATCTAGTAATGTA 180  
Qy 181 atcaataaaagaaataatctctgggaaaaaagaagaatcaacgtataaagaataaaa 240  
Db 181 ATCAATAAAAGAAAAATCTCTGGGAAAAAAGAAAGAAATCAACGTTAAATAAGAAATAAAA 240  
Qy 241 ggaattcaaaataagatatcttgaagaaataagaatacatttacaagaacagagag 300  
Db 241 GGAATTCAAATTAAGATATCTTGAGAGAAATAGATATCAATTTACAAAAGCAACGAGAG 300  
Qy 301 aaaaattttacagatgaaggacacagctatttaagatgggcatcaagggttctccagcag 360  
Db 301 AAAAAATTTACAGATGAAGGACACAGCTATTTAAGATGGGCATCAAGGTTCTCCAGCAG 360  
Qy 361 tctaaagcacaacaaagaaagagcctacctactcttttgcacaaagcagctgacatg 420  
Db 361 TCTAAAGCCAAAACAAAAGAAAGAGCCTACCTACTCTTTTGGCAAAAGCAGCTGACATG 420  
Qy 421 ggaacttgaaagctatgagaaatgctgcagcttgcctatttggaaatttggcgtg 480

Db 421 GGAACACTTGAAGACTATGGAGAAAAATGGCTGACGCTTGTCTATTGTGGAAATTTTGGCGTG 480  
Qy 481 caaataataacagcagctatccaattatagctcttgctgctaaagaagatcatgtaaa 540  
Db 481 CAAATATTAACAGCAGCTATCCAATTATATGASTCTTGGCTAAAGAAGGATCATGTAAA 540  
Qy 541 gcccaaacgcattaggattttgtctcttatttggaatagggaatggaatatgatcaagct 600  
Db 541 GCCCAAAACGCATTAGGATTTTGTCTTATGGAATAGGAATGGAATATGATCAAGCT 600  
Qy 601 aaggcactgatataattacacctttggaagctgctggaggaacacatgatgtcccagatgatt 660  
Db 601 AAGGCACATGATATATTACACCTTTTGAAGTGTGAGGAAACATGATGTCCAGATGATT 660  
Qy 661 ttggggtacagatatgttcgggaatcaatgtctcacagaattgtgaagtgtccctaaagt 720  
Db 661 TTGGGTTACAGATATTGTTCGGGAATCAATGTTCTTACAGAATTGTGAAGTTGCCCTAAGT 720  
Qy 721 tattacaagaagtgccagattattgtctgcacacatttgaaaaaagtgaaagtggttcca 780  
Db 721 TATTACAAGAAAGTGCAGATTATTTGCTGACACATTTGAAAAAAGTGAAGTGTGTCCA 780  
Qy 781 gtggaagaagtgagactaacggaagaccctgaaatctgagttctaaacagtgcagatttg 840  
Db 781 GTGGAAGTGCAGACTAACGGAAAGACCTGAAAAATCTGAGTTCTAACAGTGCAGATTG 840  
Qy 841 gattggacatatcaactactataaattttggcagaagagagagtggttcagatacaa 900  
Db 841 GATTGGGACATATACCAATCTATTAATTTTGGCAGAAAGAGAGAGTGTTCAGATACAA 900  
Qy 901 gtctctcttgacaattacatctaattgtgcaggaagaagctctagatcaggattactacaaa 960  
Db 901 GTCCTCTTGGACAATTACATCTAATTGGCAGAAAGGCTTAGATCAGGATTACTACAAA 960  
Qy 961 gcaatacactacttctaaaggcagcaagcgccggagtgcaaatgcacatgcatttata 1020  
Db 961 GCATTACACTACTTCTTAAAGGCACAAAGCCGGAGTGCAAAATGCCATTTTATTA 1020  
Qy 1021 gaaagatgtatttagaggggaatgctgcgctgcgcaaaaataacgctactactcctcaag 1080  
Db 1021 GGAAGATGTATTAGAGGGGAATGCTGCCGTGCCGCAAAAATAACGCTACTGCTTCAAG 1080  
Qy 1081 tactttccatggcagccagtaagggcaatgcaatcgcccttcattgggcttggtctctt 1140  
Db 1081 TACTTTTCCATGCAGCCAGTAAGGCAATGCAATCGGCTTCATGGGCTTGGTCTCTT 1140  
Qy 1141 tactttcatgaaaagagaggttccctctgaattatgcccraagcaactaaatactttcagaaa 1200  
Db 1141 TACTTTTCATGGAAGAGGAGTTCCTCCTGAATTATGCCRAAGCACTTAAATACTTTT 1200  
Qy 1201 gctcggaagaggggtggccgcagcagagttccagttaggcttcattgactactctgac 1260  
Db 1201 GCTGGGAAAAAGGTTGGCCCGACGACAGTTCAGATTAGGCTTACGTACTACTCTGCG 1260  
Qy 1261 tctggaatggaaggaattataaaacttgccttcaaatatttttacctggcatctcagagt 1320  
Db 1261 TCTGGAATATGAAGGATTATAAACTTGCCTTCAATATTTTACTTGGCATCTCAGAGT 1320  
Qy 1321 gggcagccctcgccatttattctgccaagatgtatgcacagggaacagagtagta 1380  
Db 1321 GGGCAGCCCTCGCCATTATTATCTGGCCAAGATGTATGCAACAGGAAACAGGAGTAGTA 1380  
Qy 1381 agatcatgcagaactgctgtggag 1404  
Db 1381 AGATCATGCAGAACTGCTGTGGAG 1404

RESULT 8  
AX148139  
LOCUS AX148139 1257 bp DNA PAT 08-JUN-2001  
DEFINITION Sequence 7 from Patent WO0136636.  
ACCESSION AX148139  
VERSION AX148139.1 GI:14347062

KEYWORDS  
SOURCE

human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1257)  
AUTHORS Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and Sands,A.T.  
TITLE Human notch ligand proteins and polynucleotides encoding the same  
JOURNAL Patent: WO 0136636-A 7 25-MAY-2001;  
Lexicon Genetics Incorporated (US)  
FEATURES  
Location/Qualifiers  
source  
1..1257  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 464 a 206 c 270 g 316 t 1 others  
ORIGIN

Query Match 60.6%; Score 1254.6; DB 6; Length 1257;  
Best Local Similarity 100.0%; Pred. No. 9.4e-239;  
Matches 1255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaagccctgtctcttaatacagattgataattcttgggtcacaattaaaact 60  
|||||  
Db 1 atgaagccctgtctctgttaatacagattgataattcttgggtcacaattaaaact 60  
|||||  
Qy 61 atcaagcagaggaacataataaagacaaaggaagaaatgtcaccacacaggtatca 120  
|||||  
Db 61 atcaagcagaggaacataataaagacaaaggaagaaatgtcaccacacaggtatca 120  
|||||  
Qy 121 gtgaacgaatcaacaatatttatcacacatatgttggaaacaaagaaacatctagtta 180  
|||||  
Db 121 gtgaacgaatcaacaatatttatcacacatatgttggaaacaaagaaacatctagtta 180  
|||||  
Qy 181 atcaataaagaaatactctcagaaaaaagaaagaaatcaacgtataaataaagaa 240  
|||||  
Db 181 atcaataaagaaatactctcagaaaaaagaaagaaatcaacgtataaataaagaa 240  
|||||  
Qy 241 ggaattcaaaataaagatatcttgaagagaataagaatcatttcacaaagcaagcagag 300  
|||||  
Db 241 ggaattcaaaataaagatatcttgaagagaataagaatcatttcacaaagcaagcagag 300  
|||||  
Qy 301 aaaaattttacagatgaaggagacagctatttaagatgggcatcaaggtttccagcag 360  
|||||  
Db 301 aaaaattttacagatgaaggagacagctatttaagatgggcatcaaggtttccagcag 360  
|||||  
Qy 361 tctaaagccaaacaaacaaagaaagccctacctcttcttggcacaagcaagcagatg 420  
|||||  
Db 361 tctaaagccaaacaaacaaagaaagccctacctcttcttggcacaagcaagcagatg 420  
|||||  
Qy 421 ggaacttgaagctatggagaaaatggctgacgcttctgctatttggaaaatttggcgtg 480  
|||||  
Db 421 ggaacttgaagctatggagaaaatggctgacgcttctgctatttggaaaatttggcgtg 480  
|||||  
Qy 481 caaataataacgcagctatccaattatatagctccttggcttgcttgaaagaaggatcatgt 540  
|||||  
Db 481 caaataataacgcagctatccaattatatagctccttggcttgcttgaaagaaggatcatgt 540  
|||||  
Qy 541 gcccaaacgattagattttgtctctcttgaatgaatgaatgaatgaatgaatgaatgaat 600  
|||||  
Db 541 gcccaaacgattagattttgtctctcttgaatgaatgaatgaatgaatgaatgaatgaat 600  
|||||  
Qy 601 aaggcactgatattacacctttggaagctgtgaggaaacatgatgtcccagatgatt 660  
|||||  
Db 601 aaggcactgatattacacctttggaagctgtgaggaaacatgatgtcccagatgatt 660  
|||||  
Qy 661 ttgggggtacagatatttgcgggaatcaatgtttctacagaattgtgaagttgccctaaagt 720  
|||||  
Db 661 ttgggggtacagatatttgcgggaatcaatgtttctacagaattgtgaagttgccctaaagt 720  
|||||  
Qy 721 tattacaagaaagtgccagatttatattgtcgcacatttgaacatttgaagtggaagttcca 780  
|||||  
Db 721 tattacaagaaagtgccagatttatattgtcgcacatttgaacatttgaagtggaagttcca 780  
|||||

Qy 781 gtggaaaagtggagactaacggaagacctgaaactctgagttcttaacagtgagattttg 840  
|||||  
Db 781 gtggaaaagtggagactaacggaagacctgaaactctgagttcttaacagtgagattttg 840  
|||||  
Qy 841 gattgggacataaccatactataataatttttggcagaaaggagagtgctcagatacaa 900  
|||||  
Db 841 gattgggacataaccatactataataatttttggcagaaaggagagtgctcagatacaa 900  
|||||  
Qy 901 gtctcttggacaattacatctaattgacaggaaggtcctagatcagattactacaaa 960  
|||||  
Db 901 gtctcttggacaattacatctaattgacaggaaggtcctagatcagattactacaaa 960  
|||||  
Qy 961 gattacactacttcttaaggcagcaaaaggccgggagtgcaaatgccatggcatttata 1020  
|||||  
Db 961 gattacactacttctttaaaggcagcaaaaggccgggagtgcaaatgccatggcatttata 1020  
|||||  
Qy 1021 ggaagaatgatttagaggggaatgctgcccgtccgcgcaaaaataacgctactgcctcaag 1080  
|||||  
Db 1021 ggaagaatgatttagaggggaatgctgcccgtccgcgcaaaaataacgctactgcctcaag 1080  
|||||  
Qy 1081 tactttccatggcagccagtaaggcgaatgaatcgcccttcattggcttgctctt 1140  
|||||  
Db 1081 tactttccatggcagccagtaaggcgaatgaatcgcccttcattggcttgctctt 1140  
|||||  
Qy 1141 tactttcatggaaaaggagttccctgaattatgcccraagcaacttaatactttcagaaa 1200  
|||||  
Db 1141 tactttcatggaaaaggagttccctgaattatgcccraagcaacttaatactttcagaaa 1200  
|||||  
Qy 1201 gctgcggaagaagggtggccgcagcagctccagttccagttcagttcagttact 1255  
|||||  
Db 1201 gctgcggaagaagggtggccgcagcagctccagttccagttcagttcagttact 1255  
|||||

RESULT 9

LOCUS E26632 2382 bp DNA PAT 07-FEB-2001  
DEFINITION TSA305 gene.  
ACCESSION E26632  
VERSION E26632.1 GI:13018167  
KEYWORDS JP 199215987-A/1.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 2382)  
AUTHORS Yosuke,H.K.O.O.  
TITLE TSA305 gene  
JOURNAL Patent: JP 199215987-A 1 10-AUG-1999;  
COMMENT OTSUKA PHARMACEUT CO LTD  
OS Unidentified  
PN JP 199215987-A/1  
PD 10-AUG-1999  
PF 20-APR-1998 JP 1998136803  
PR

PI YOSUKE HARADA,KOICHI OZAKI  
PC C12N15/09,C07K14/47,C07K16/18//C07K14/435,C12N5/10,C12Q1/68,  
PC C12N15/00,  
PC C12N5/00  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..2382 /organism='Unidentified'.  
FEATURES  
Location/Qualifiers  
1..2382  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 689 a 501 c 614 g 578 t  
ORIGIN

Query Match 30.5%; Score 631.4; DB 6; Length 2382;  
Best Local Similarity 60.5%; Pred. No. 2.6e-115;

**Matches** 1056; **Conservative** 0; **Mismatches** 686; **Indels** 3; **Gaps** 1;

Qy	285	acaaagcaagcagagaaaaattttacagatgaagagagaccagcgtatttaagatggtgcat	344
Db	513	AGAAGAGGCTGCTTAAGAGAGCGCAGATGCAGGAAGCAAAATGATGATCAAACTGGAAAT	572
Qy	345	caagggtctccagcagctctaaagccaaacaaagaaagaaagcctaccactcttttgc	404
Db	573	GAAATYCCCTTAATGGAGGCAATTAAGAAAAGCCAAAAGAGAAAGCAATATCGGTATCTCCA	632
Qy	405	caagcagctgacatgggaaacttgaaagctatggagaaaaatggctgacgcctttgctatt	464
Db	633	AAAGGCAGCAAGCATCAACCATACCAAAGCCCTGGAGAGAGTGCATATGCTCTCTTTATT	692
Qy	465	tggaaaattttggcgtgcaaaaataaacacagcagctatccaattatagtccttggcctaa	524
Db	693	TGGTGATTAATCTGCCACAGAAATATCCAGACGCGAGAGAGATGTTTGAGAAAGCTGACTGA	752
Qy	525	agaagatcatgtaaaagcccaaacgcattagatttttctctctatggaataggaaat	584
Db	753	GGAAAGGCTCTCCCAAGGACAGACTGCTCTTGGCTTCTCTGTATGCCCTCGGACTTGGTGT	812
Qy	585	ggaatatgtataagactaaaggccactgatattataccacttttggaaagtgcctggaggaacaat	644
Db	813	TAATTCAAAGTCAGGCAAAAGGCTCTTGATATATATATATACATTTTGGAGCTCTTGGGGCAATCT	872
Qy	645	gaigtccagatgattttgggtacagatattgttcgggaatcaaatgttctacagaattg	704
Db	873	ANTAGCCCACTGGTTTGGGTACAGATACTTGGGCTGGCATCGGCTCTCCAGAGTTG	932
Qy	705	tgaagtgcctaagtattacaagaaaagtggcagattattctgacacatttgaaaa	764
Db	933	TGAATCTGCCCTGACTACTATCGTCTTGTGGCCAAATCATGTTGCTAGTGATATCTCGCT	992
Qy	765	aagtgaaggtgtccagtgaaaagtgagactaacgaaagacacctgaaacttgagttc	824
Db	993	AACAGGAGGCTCAGTAGTACAGAGAAATACGGCTGCCTGATGAAGTGAAATAATCCAGGAAT	1052
Qy	825	tacagttgagattttggattgggacataaccaaactataaaatttttggcagaagaagg	884
Db	1053	GAACAGTGGAAATGCTAGAGAAGATTGTGATTCAAATATTACCAGTTCTTGTAGCTGAAAAAGG	1112
Qy	885	agatgttcagatacaagtcctctcttgacaaatcacatctaatttggcaggaaaagctctaga	944
Db	1113	TGATGTACAAGCAGAGGTTGGTCTTTGGACAACCTGCACCTGCACGGAGGGCGTGGAGTAGA	1172
Qy	945	tcaggtattactacaagacattacactactctttaaaggcagcaagcgcgggagtgcaaa	1004
Db	1173	ACAGAATCATCAGAGAGCAATTTGACTACTTCAATTTTACGACCAAAATGCTGGCAATTCACA	1232
Qy	1005	tgccatggcatttataggaagaatgatttagaggggaatgtccgtgcgcgcaaaataa	1064
Db	1233	TGCCATGGCCCTTTTGGGAAGAAGTGTATTTCGGAAGGAAGTGACATTTGTACCTCAGAGTAA	1292
Qy	1065	cgctactgcctcaagtaacttttccatggcagccagtaagggccaatgcaatcgccctca	1124
Db	1293	TGAGACAGCTCTCCACTACTTTTAAAGAAAGCTGCTGACATGGGCAACCCAGTTGGACAGAG	1352
Qy	1125	tgggcttggctcttcttacttttcagaaaaggattccccctggaattatgcgccaagcaat	1184
Db	1353	TGGGCTTGGAAATGGSCCTACCTCTATGGGAGAGAGTTCAAAGTTAATATGATCTAGCCCT	1412
Qy	1185	taaaactttcagaagctgcggaaaaaagggtggtcccgacgcacagtttccagttagtcctt	1244
Db	1413	TAAATATTTCCAGAAAGCTGCTGAACAAGGCTGGGTGGATGGGAGAGCTACAGCTTGGTTC	1472
Qy	1245	catgtactactctggctctggaatatggaagattataaaacttgccttccaaattttta	1304
Db	1473	CATGTACTATTAATGGCAATTTGGAGCTCAAGAGAGATTATTAACAGGCCCTTTGAAGTATTTTAA	1532
Qy	1305	cctggcaactccagatggggcagccctcgccacttttatctgtgccaaagtatgatgcaac	1364
Db	1533	TTTAGCTCTFCAGGAGGCCATATCTTGGCTTCTTATTAACCTAGCTCAGATGCAATGTCGAG	1592

[illegible]

ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 3473)
AUTHORS	Biunno, I., Cattaneo, M., Volorio, S. and Zollo, M.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAR-1998) ITBA, CNR, Via Fratelli Cervi 93, Segrate MI 20090, Italy
FEATURES	Location/Qualifiers
source	1. .3473 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="14" /map="14q31" /dev_stage="adult"









```
QY 1785 ttatatgtatgaacagcgttaggcatacacaagaagacattcacttgccagaaagtgtga 1844
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2058 ATATATGTCATGAAGAGGCTGGGCATTAACAGGATATTCACCTTGCGCAAAAGCTTTTA 2117

QY 1845 cgaatgctgtctaaacagatccagatgcccacatacactgctgtctctttggcgtcatga 1904
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2118 TGACATGGCAGCTGAAGCAGCCAGATGACAAAGTTCAGTCTTCCTAGCCCTCTGCAA 2177

QY 1905 actggaatactgcattgctcgggatactcgtgtttttaatcagtttcacacacagatg 1964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2178 ATTGGCGTCTGCTATTTCTTCAGTACATACGG---GAAACAACATTCGAGATATGTT 2234

QY 1965 gaactggctgaactggaacacacacattggaccacactgggacttatttggattggcct 2024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2235 CACCAAACTGTATATGGACCACTTTTGGGACCTGAGTGGGACCTTTACCTCATGACCAT 2294

QY 2025 cattg 2029
|||||
Db 2295 CATTG 2299

RESULT 12
LOCUS AB020335 7885 bp mRNA PRI 04-DEC-1999
DEFINITION Homo sapiens Pancreas-specific TSA305 mRNA, complete cds.
ACCESSION AB020335
VERSION AB020335.1 GI:6518494
KEYWORDS TSA305.
SOURCE Homo sapiens pancreas cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 7885)
AUTHORS Harada,Y.
TITLE Pancreas-specific gene
JOURNAL Published Only in Database (1999) In press
REFERENCE 2 (bases 1 to 7885)
AUTHORS Harada,Y.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-1998) to the DDBJ/EMBL/GenBank databases. Yosuke
Harada, Otsuka Pharmaceutical Co., Ltd., OTSUKA GEN Research
Institute, 463-10 Kagasuno Kawauchi-cho, Tokushima, Kagasuno,
Kawauchi-cho, Tokushima 771-0192, Japan (E-mail:harada@otsuka.gr.jp,
Tel:81-886-65-2888, Fax:81-886-37-1035)
FEATURES
Location/Qualifiers
1..7885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3-31.1"
/tissue_type="pancreas"
46..2430
/gene="TSA305"
46..2430
/gene="TSA305"
/notes="Pancreas-specific gene"
/codon_start=1
/protein_id="BAA87904.1"
/db_xref="GI:6518495"
/translation="MKRVRLGLTLCLLCVLLSLASSEESQDSLSKSTLTLSDES
VKDHTAGRVVAGQIFLDSSESELESSELSQESGSESVTEIDIFLESPPNPK
DYEFKVKYKRALTAIEGTAHGEPCHPFLFDKEYDECTSDGREDRLWCATTDYK
ADEKMGFCETEAAKRMQEAEMMYOTGMKILGNSKQKREARYFLOKAAMNHN
TKALERVSYALLFGDYLQNTQAAREMEFKLTEGSPKQOTALFLYASGLVNSOA
KALVYFTGALGNLIAMVLGYRWAGICVLQSCESALTHRLVANHVASDISITGG
SVVQRIRLPDENPMNSGMLLEDLIQYFLAEKGDVQAOVLHGLQHLHGGNVEQ
NHRQAFVFLNAANAGNSHAMAFGLKMTYSEGSDIVPQSNETHLYFKKAADMGVPVQ
SLGMAVLYETREAAKRYDILALKYFOKAAEQGVNDGQLQGLGSMYNGICGVKRYKQALK
YFNLASOGGHILAFYNLAOMHAGSTGVNRSCHTAVLEFKNVCERWSERLMTAYNSY
KGDYNAIVIOYLLAEQGYEVAOSNAFILLDOREASIVGENETYPALLHNRASQ
GYTVARIKGLDHYFFGTDVDYETAFTHYRLASSEQHSQAQMFNLGMHKEGIGIKQ
DIHLAKRFYDMAAEASPAQVPVFLALCKLGVVYFLQVIRETNIRDMFTQLDMDQLLG
```

```
BASE COUNT 2332 a 1466 c 1619 g 2468 t
ORIGIN
PEWDLYLMTIIALLGLTVIAYRQRHQHDMAPRPFPAPPPQEGPBPQBPQ"
Query Match 30.5%; Score 631.4; DB 9; Length 7885;
Best Local Similarity 60.5%; Pred. No. 2.5e-115;
Matches 1056; Conservative 0; Mismatches 686; Indels 3; Gaps 1;
QY 285 acaaaagcaagcagagaaaaattttacagatgaagagaccagctatttaagatggcat 344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 558 AGAAGAGGCTGCTTAAGAGCGGAGATGCGAGGAAGCAGAAATGATGTATCAAACTGGAAT 617

QY 345 caaggttctccagcagctctaaaagccaaaacaaagaaagcctacactcttttgc 404
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 GAAATCCCTTAATGGAAGCAATAAGAAAGCCAAAGAGAGCATATCGGTATCTCCA 677

QY 405 caaagcagctgacatgggaaacttgaaagctatgggaaaaatggctgaogcttgcatt 464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 678 AAAGGCAGCAAGCATGAACCATACCAAGCCCTGGAGAGAGTGTATATCTCTTTTATT 737

QY 465 tggaaattttggcgtgcaaaataaacaacagcagctccaattatagctccttgctaa 524
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 738 TGTGATTACTTGCACAGAAATATCCAGGCGAGAGAGATGTTTCAGAAAGCTGACTGA 797

QY 525 agaagcatcatgtaaaagcccaaaacacgcatcttaggattttgtcttcttcttatgaaat 584
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 798 GGAAGGCTCTCCCAAGGACAGACTGCTCTTGGCTTCTGTATGCCCTCTGGACTTGGTGT 857

QY 585 ggaatatgatcaagctaaagcactgatataattacacctttgggaagtgcctggagaacat 644
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 858 TAATTCAAGTTCAGGCAAGGCTCTTGTATATTATACATTTGGAGCTCTTGGGCAATCT 917

QY 645 gatgtcccagatgattttgggtcacagatatgttcgggaatcaatgttctcagaattg 704
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 918 AATAGCCCAATGGTTTGGGTACAGATACTGGGCTGGCATCGGCTCTCCAGAGTTG 977

QY 705 tgaagttgcctcaagtatttacagaagtgagcagattattgtcgacacatttgaaaa 764
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 978 TGAATCTGCCCTGACTCACTATCGTCTGTTGTCGAATCATGTTGCTAGTCATATCGCT 1037

QY 765 agtgaagtggttccagtggaagaaagtgaactaacaggaaagacctgaaactgtgagttc 824
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1038 AACAGAGGCTCAGTAGTAGTACAGAGAAATACGGCTGCCCTGTATGAAGTGAATAACAGGAAT 1097

QY 825 taacagtgaattttgattggagacataataccaataactataaatttttggcaaaagag 884
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1098 GAACAGTGAATGCTAGAAGAAGATTTGATTCATATATTACAGTTCTCTAGCTGAAAAGG 1157

QY 885 agatgttcagatacaaaagtctctcttggaacaattacatcttaattggcaggaaaggtctaga 944
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1158 TGATGTACAAGCACAGGTTGGTCTTGGACAACCTGCACCTGCACGGGGCGTGGAGTAGA 1217

QY 945 taagattactacaagcattacactactcttcttaagggaagcgaagccggggagtgcaaa 1004
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1218 ACAGAATCATCAGAGAGCATTTTGCATCTACTTCAATTTAGCAGCAAAATGCTGGCAATTCACA 1277

QY 1005 tgcctatggcatttatagaaaagatgtatttagagggaatgctgcctgctgcgcaaaataa 1064
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1278 TGCATATGGCTTTTGGGAAAAGATGATTTCGGGAAGGAAGTGACATTTGTACCTCAGAGTAA 1337

QY 1065 cgtactgctctcaagtacttttcccttgagcagcagtaaggccaatgcaatcgcccttca 1124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1338 TGAGACAGCTCTCCACTACTTTTAAAGAAGCTGCTGACATGGGCAACCCAGTTGGACAGAG 1397

QY 1125 tgggcttggtctcttactttctaatgaaagagttccctgaattatgcccraagcact 1184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1398 TGGGCTTGGAAATGGCCCTACTCTTATGGAGAGAGTGAAGTTAAATATATGATGATGACCT 1457

QY 1185 taactacttcagaaagctgcggaagaaaggggtggccgacgacacattccagttagctt 1244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1458 TAAGTATTTCCAGAAAGCTGCTGAACAAGGCTGGGTGGATGGGCAGCTACAGCTTGTTTC 1517
```

```
QY 1245 catgtactactgtgctctggaatggaagattataaaacttgccttcaaatatttta 1304
Db 1518 CATGTACTATAATGGCATTTGGAGTCAAGAGAGATTATAAACAGGCTTGAAGTTTAA 1577
QY 1305 cctggcatctagatggcgagccctcgcattttattctctgcccagaatgatgcaac 1364
Db 1578 TTTAGCTTCTCAGGAGGCCATATCTTGCTTTCTATACCTAGCTCAGATCAATGCCAG 1637
QY 1365 aggaacagagtagtaaagatcatgcagaaactctgtgagctcttataaagggtgctgtga 1424
Db 1638 TGGCACCAGCGTGATGCGATCATGTCACACTGCAGTGGAGTTGTTTAAAGATGTATGTA 1697
QY 1425 actaggccaactgggtcgagaaattcctgcagacttacttggctataaagatgggtgat 1484
Db 1698 ACGAGGCCCTTGGTCTGAAGGCTTATGACTGCTTATACAGCTATAAAGATGGCGATTA 1757
QY 1485 agattctctctgttcagatgcactgcttgcagaaatggggtatgaagtgcctcaaa 1544
Db 1758 CAATGCTGCAGTGATCCAGTACTCTCTCTGCTGAACAGGGCTATGAAGTGGCACAAG 1817
QY 1545 caaticcagcattcattttggaatctaaaaaggctaacattcttgaaaaagagaagtga 1604
Db 1818 CAATGCAGGCTTTATCTTGATCAGAGAGAACAGCATTTGAGTGAGAACTGA 1877
QY 1605 tocaatggcgctctctctatggaatcgagctgccattccaagcaatgcatttgcctagat 1664
Db 1878 TCCAGAGCTTTGCTTACATTTGGAACAGGGCGCTCTCAAGGCTATACCTGTGCTAGAT 1937
QY 1665 aaaaattggagattaccattactatggctatgggactaagaagactatcaaacagcagc 1724
Db 1938 TAAGCTCGAGAGACTACCAATTTCTATGGGTTTGGCCAGGATGTAGATTATGAACCTGCATT 1997
QY 1725 cacacactacagcattgcagcacaacaaataccacaacgcgcagccatttcaactctgc 1784
Db 1998 TATTCTATACCTCTGGCTTCTGACGAGCAACACAGTGCAGCAAGCTATTTTAACTGGG 2057
QY 1785 ttatatgtatgaacacggttaggcatacaaaagacattccttggccagaagattga 1844
Db 2058 ATATATGATGATAAGAGGACTGGGCATTTAAACAGGATATTCACCTTGGCAACGTTTAA 2117
QY 1845 cgacatggctgtcnaaagatccagatccacatacctgtgcttcttggcgcgtcatga 1904
Db 2118 TGACATGGCAGCTGAAGCGAGCCAGATGACAAAGTTCAGTCTTCCCTAGCCCTCTGCAA 2177
QY 1905 actggaatacagcattgctgcggatattcctgttttataatcagttcacacacagatg 1964
Db 2178 ATTGGCGTCTCTATTCTTCGAGTACATACGG---GAAACAACATTCGAGATATGTT 2234
QY 1965 gaactgggtgaactggacaacacacattgggaccacactggggacttatttggattggcct 2024
Db 2235 CACCAACTTGATATGGACCACTTTTGGGACCTGAGTGGGACCTTACCTCATGACCAT 2294
QY 2025 cattg 2029
Db 2295 CATTG 2299

RESULT 13
AB048195
LOCUS AB048195 2441 bp mRNA ROD 02-SEP-2000
DEFINITION Mesocricetus auratus mRNA for SEL1L, complete cds.
ACCESSION AB048195
VERSION AB048195.1 GI:9967439
KEYWORDS Mesocricetus
SOURCE Mesocricetus auratus cell_line:Hamster kidney BHK cDNA to mRNA.
ORGANISM Mesocricetus auratus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Mesocricetus.
1 (sites)
REFERENCE:
AUTHORS Wada, M. and Moriyama, T.
TITLE Hamster kidney SEL1L
JOURNAL Published Only in Database (2000) In press
```

## REFERENCE

2 (bases 1 to 2441)  
Wada, M. and Moriyama, T.  
Direct Submission  
Submitted (31-AUG-2000) to the DDBJ/EMBL/GenBank databases. Tatsuya Moriyama, Kyoto University, Research Institute for Food Science; Gokasho, Uji, Kyoto 611-0011, Japan  
(E-mail: moriyama@soya.food.kyoto-u.ac.jp, Tel: 81-774-38-3753, Fax: 81-774-38-3752)

## FEATURES

source

1. 2441

/organism="Mesocricetus auratus"

/db\_xref="taxon:10036"

/cell\_line="Hamster kidney BHK"

13. 2397

/gene="SEL1L"

13. 2397

/gene="SEL1L"

/note="similar to Caenorhabditis elegans sel-1"

/codon\_start=1

/product="SEL1L"

/protein\_id="BAB12403.1"

/db\_xref="GI:9967440"

/translation="MQVHVGLTLCLCAVLLSATSSDSDNQSDISIDSKSLPADES  
VKDSTTRVWAGQIFVDSDEPEVESPLQEEESKTEEVSVGDEI3FVESPSSSK  
SYEAKRKRPVLTAEHTAHGEPHFPLFLDKYEDCTSDKREARYVLOKAAAMNH  
TDEKWFCTEEDAAKRMOEAEMIYQAGMKILNGSKRSOKREARYVLOKAAAMNH  
TKALERSVALLFGDYLNTIOQAAMEFEKTEEGSPKQATGALFLYVSGLGVSNOA  
KALVYTFGALGGLLIAMVLGYRWAGIGVLQSCSVLTHYRLVANHVADISLITGG  
SVVORILPDEVENPGMNSGMEEDLIQYFQAEKGDVQAGVGLQGLHLHGGRVEQ  
NHRADYFNLAAGNAGNSHAFLEKMYSEGSDIVPQSNETHALHYFKAAADGNVPVQ  
SGLMAYLYGRGIOVNDLALYFQAAEQGWVDGLOLQSGMYNGIGVKRDYKQALK  
YFNLASOGGHILAFYLAQMSAGTGVNRSCHTAVLEFKNVCERGRWELMTATNSY  
KGDYNAAVOYLLLAEGYVVAOSNAFILDREATVIGENETVPRALLHNRASO  
GYTVARKLDYHFFYFGTDYDYEAFIHYRLASEQOHSQAAMFNGLYHHEKGLGIKO  
DIHLAKRFYDMAAESPDQVPFLALCKLGVVYFLQYIIRANIRIDIFETLQDMDLLG  
PEWDLYMTIIALLGLTVIAYRQRHQDQVPPRPPEPWPAPEQEGPEQPPQPPQ"

BASE COUNT 649 a 579 c 678 g 535 t

## ORIGIN

Query Match 29.6%; Score 612.2; DB 10; Length 2441;  
Best Local Similarity 59.8%; Pred. No. 1.7e-111;  
Matches 1044; Conservative 0; Mismatches 698; Indels 3; Gaps 1;

QY 285 acaaaagcaagcagagaaaaattttacacatgaagagaccagctatttaagatggcat 344  
Db 525 AGAAGATGCTGCTCAAAAGACGGCAGATGCAGGAAGACAGAAATGATTTACGAGCTGGGAT 584  
QY 345 caaggtctccagcagctctaaaagccaaaacaaaaagaagagcctacctcttttgc 404  
Db 585 GAAGATACCTGATGGGAGCAACAGAAAGACCCAAAGAGAGAGCATATCGGTATCTTCA 644  
QY 405 caaagcagctgacatgggaaccttgaaagctatggagaaaaatggctgacgcttggctatt 464  
Db 645 AAAGGCAGCAGGCATGAATCACACCAAGCCCTGGAGAGAGATCTCTACGCTCTCTTATT 704  
QY 465 tggaaattttggcgtgcataataataaacagcagctatccaatatatagctcttggctaa 524  
Db 705 TGGTGATTACCTCACACAGAAATATCCAGCGGCTTCAAGAGATGTTTGAAGAACTGACCCA 764  
QY 525 aagaagcatcatgtaagcccaaaacgcattaggttttcttcttcttatgggaataggaat 584  
Db 765 GGAAGGTCTCCCAAGGACAAACTGCTCTTGGCTTCTACGCTCTGGAGCTCGGTGT 824  
QY 585 ggaatatgatcaagctaaagcagcactgatataattacaccttttgaagtgtctgggaagaaat 644  
Db 825 TAATTCAAGTCAGGCAAGGCTCTTGTATATTATATCTTTTGGAGCCCTTGGAGGCAATCT 884  
QY 645 gatgtccacagatgatttggggtacagatatttgcgggaatacaatgttctctcagaaatg 704  
Db 885 GATAGCCCATATGGTTTGGGTACTCCGGTACTGGGCTGGCAGTCCGAGTCTCGAGAGTTG 944  
QY 705 tgaagtgcctcaagtattatacaagaagtggcagattatatattgctgacacatttgaaaa 764

Db 945 TGAGTGGTACTGACCAATATCGTCTGTTGCCAACCAATGTTGTAGCATATCCCT 1004  
Qy 765 aagtaagtggtccagtggaagagtgagactaaacgaaagacctgaaatctgagttc 824  
Db 1005 GACCGAGGCTCAGTAGTCCAGAGAAATACGGGTGCTGATGAAGTGGAAACCCAGGAT 1064  
Qy 825 taacagtgagatttggtgagacataataactataataatttttggcagaagaaggg 884  
Db 1065 GAACAGTGGGTGGAAGAGGATTTGATTCCAGTATTACCAAGTTCCCTGGCTGAGAAGGG 1124  
Qy 895 agatgtcagatacaagtcctctctggacaattacatactaatggcaggaaggtctaga 944  
Db 1125 TGACGTGAAGCACAGGTTGGTCTGGGCAACTGCATCTGCATGAGGCGGTGGAGTAGA 1184  
Qy 945 tcaggtactactacaagcatcacactactcttctaaagcagcaaaagccggagtgcaaa 1004  
Db 1185 ACAGAAATCACAGAGAGCATTTGACTACTTCAACTTACAGCAAAATGTCGCAATTCACA 1244  
Qy 1005 tgcctggaattataggaagatgtattagagggggaatgctgcgtgcccgaataaa 1064  
Db 1245 TCGCATGGCTTCCTGGGAAGATGATTCTGAAGGAAGCGACATCGTCCCTCAGAGTAA 1304  
Qy 1065 cgtactgcttcaagtaacttttccatggcagccagtaaggcaatgcaatcggcctca 1124  
Db 1305 TGAGACAGCACTTCACTACTTCAAGAAAGCCGCGACATGGGCAACCCCGTGGGACAGAG 1364  
Qy 1125 tgggttggtctcttactttcattggaagagggttcccttgaaattatgcccraagcaat 1184  
Db 1365 CGGGCTTGGAAATGGCTACTCTATGGAAGGGGAATTCAGTTAATATGACTGGCACT 1424  
Qy 1185 taaactctcagaagctgcggaaaaaggggtggccagcagcacagttccagttaggctt 1244  
Db 1425 CAAGTATTTCCAGAAAGCTGTGACAAAGGCTGGTGGATGGCAGCTGCAGCTTGGCTC 1484  
Qy 1245 catgtactactcgtctggaatgaatgaagagattataaactgcttcaaatatttta 1304  
Db 1485 TATGTACTACAAATGGCATTTGGAGTCAAGAGAGATTATAAGCAGGCTTGAAGTATTTAA 1544  
Qy 1305 cctggcatctcagtgaggcagccctgcctatttatctctgcccagcagcatgatgcaac 1364  
Db 1545 TTTACATCTCAAGAGGCCATATCTGTGCTTCTATACCTTGGCAGCATGCCAGCCAG 1604  
Qy 1365 aggaacagagtagtaagatcatcagaaactcgtgagctgtttataaagggtctgtga 1424  
Db 1605 CGGCACAGGGGTGATGAGTCTGTGCACACTCGGTGGAGTGTGTTAAGAACGTGTGTA 1664  
Qy 1425 actaggccaactgggtgagaaattcctgcagcagcttacttgcctataaggatgatat 1484  
Db 1665 GCGAGGCCGTGGTGGGAGGCTCATGACTGCCCTACACAGCTATAAGACGGGGACTA 1724  
Qy 1485 agattctctctgttcagtagtgcactgcttcagaaaaatggggtgatgaagtgcctcaa 1544  
Db 1725 CAATGCTGCAGTGGTCCAGTACTCTGCTGCCCAACAGGCGCTAGGAAGTGGCAGAG 1784  
Qy 1545 caattcagcatcatttgggaatctaaaaaggctaacattcttggaaaaagagaagtga 1604  
Db 1785 CAACGACGCTTCATCCTTGACCAGAGAGAAGCAACCACTTGTAGGTGAGAAATGAGACTTA 1844  
Qy 1605 tccaatggcgtctcctcatggaatcagagctgccttcaaggaatgcaattgcttagagt 1664  
Db 1845 CCCCAGAGCGTTGCTGCATTTGGAACAGGCGCCCTCTCAAGGTTTATACCGTGGCTAGAA 1904  
Qy 1665 aaaaattggagattaccattactatggctatgggactaagaagaactatcaaacagcagc 1724  
Db 1905 TAAACTTGGAGACTACCATTTCTATGGGTTGGCACTGATGTGGATTTACGACACTGCTT 1964  
Qy 1725 cacacactacagcattgagcagaacaaataccacacgcgcaagcattgttcaatctggc 1784  
Db 1965 TATTATTACCAGCTGGCTTCCGAGCAACAACAGAGTGCACCAAGCTATGTTTAACTGGG 2024  
Qy 1785 ttatatgtatgaacacgcttaggcatacaaaagacattcacttggccagaagattga 1844

BASE COUNT 955 a 881 c 953 g 874 t

Db 2025 CTACATGATGAGAAGGGCCTAGGCATTAAACAGGATATTACCTTGCAAAACGCTTTTA 2084  
Qy 1845 cgcacatgctgctcctcaacgagtgccagatgccacataccctgtgctctttgctgcgtcatgaa 1904  
Db 2085 TGACATGGCAGCGGAGGTAGCCAGATGACAAAGTCTCTGCTCTCCCTGCACTCGCA 2144  
Qy 1905 actggaactacgcatttgcgtccggatatacctgtttttaaactagtttcacaacagatg 1964  
Db 2145 ATTGGTGTGCTCTATTCTTACAGTACATACGGAAGCAAAAT--ATCCGAGACATATT 2201  
Qy 1965 gaactggcgtgaaactggacaacacacattggaccacactggggacttatttgatggcct 2024  
Db 2202 CACCAACTGGATATGGACAGCTCTTGGAGCCGAGTGGGACCTTTACCTCATGACCAT 2261  
Qy 2025 cattg 2029  
Db 2262 CATCG 2266

RESULT 14  
AF063095  
LOCUS AF063095 3663 bp mRNA ROD 15-JAN-1999  
DEFINITION Mus musculus SELIL (Selil) mRNA, complete cds.  
ACCESSION AF063095  
VERSION AF063095.1 GI:4159994  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3663)  
AUTHORS Donoviel, D.B., Donoviel, M.S., Fan, E., Hadjantonakis, A. and Bernstein, A.  
TITLE Cloning and characterization of Sel-1l, a murine homolog of the C. elegans sel-1 gene  
JOURNAL Mech. Dev. 78 (1-2), 203-207 (1998)  
MEDLINE 99077704  
REFERENCE 2 (bases 1 to 3663)  
AUTHORS Donoviel, D.B., Donoviel, M.S., Fan, E., Hadjantonakis, A.-K. and Bernstein, A.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAY-1998) Samuel Lunenfeld Research Institute, Room 983, Mount Sinai Hospital, 600 University Avenue, Toronto, ON M5G-1X5, Canada  
FEATURES  
source 1. 3663  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1. 3663  
/gene="Selil"  
76..2298  
/gene="Selil"  
/note="similar to Caenorhabditis elegans SEL-1, negative regulator of Notch; contains putative PEST sequence"  
/codon\_start=1  
/product="SELIL"  
/protein\_id="AAD05210.1"  
/db\_xref="GI:4159995"

/translation="MOVPRVRLSLLLCVLLGSAATSDDKTNOODSLDSKSLPTDE  
SVKHTTTGKVVAQIFVDSAEAEVESLQDEEDSSKTQEEISLESFSPNSSKTYEE  
LKRVRPVLTAEAEEDAAKRMOEAEIYQAGMKILNGSNRKSQKREARYLQKAA  
GNNHTKALERSYALLFGDYLTONIAAKEMFKLEEGSPKGTGLGFLVAGLGVN  
SSOKALVYITFGALGNLIAHMLIGYRWAGIGVLQSCESALTHYRLVANHVASDIS  
LTGGSVWORIRLPDEVENPGMNSMLEEDLYQYFLAEKGDVQAOVLGOLHLHGR  
GVEONHOFADYFNLAANGNSHAMFLGKMYSEGSDIVPOSETALHYFKAADMGN  
PVGSGLGMAIYGRGVVNDLALKYFQKAAQGVQGLQSGMNYNGVGVKRDYK  
QALKYFNLAQSGGHILAFYLAQMSHAGTGVNRSCHTAVELFNKVCNERGWSERLMTA  
YNSYKDEYDINAQVQYLLALQGEVAGSNAAFILDQREATIVGNETYPRALLHNR  
AASQCYTVARIKFDYHFYFGTVDYETAFTHYLAHQHQAQMLNGLVYMHKGL  
GIKODIHLAKRFYDMAEASPDQVPVFLALCKLVGVFLOYIREANRLDFTQDMD  
QLLGPWDLYLMTIALLGVIAVQRQHDIPVPRPPRPAPPPQEQEQPPQ

## ORIGIN

Query Match 29.3%; Score 607.2; DB 10; Length 3663;  
Best Local Similarity 60.0%; Pred. No. 1.6e-110;  
Matches 1030; Conservative 0; Mismatches 683; Indels 3;

1;

QY	314	atgaaggagacgctatttaagatg9gggcatcaaggttctccagcagctctaaaaagccaaa	373
DB	455	AGGAACACAGAGATGATCTATCAGCGGGATGAAGATACTGAATGGAACGAATAGGAAGA	514
QY	374	acaaaaagaagaagcctacctacttttgcacaaagcagctgacatggaacactgaaag	433
DB	515	GCATAAGAGAGAAGCATATCGGTACCTTCAGAAGCGACAGGCATGAATCACACCAAG	574
QY	434	ctatggagaaaaatggcgctcttctgctattttggaaaattttggcgtgcaaaataaacag	493
DB	575	CCCTGGAGAGAGTGTCTTATGCTCTCTTGTGGTGATTACCTCACACAGAATATCCAGG	634
QY	494	cagctatccaattatagtccttgctgctataaagaagatcatgtaaaagcccaaaacgcac	553
DB	635	CAGCAAAGAGATGTTTGAAGAACTACTGAGGAAGGGCTCTCCCAAGGACAGACTGGTC	694
QY	554	taggatttttgcctcttatggaaataggaaatggaatgatacaagctaaggcactgat	613
DB	695	TTGGCTTCTCTACGCTTCTGGGCTTGGNGTTAAATTCAAGTCAGGCAAAAGGCTCTTGAT	754
QY	614	attacacctttggaagcgtcgagagaaacatgatgccagatgattttgggggtacagat	673
DB	755	ATTATACTTTCGGAGCTCTTGGAGCAACCTGATAGCCCATGATGATTTTGGGTTCACGCT	814
QY	674	atttgcgggaaccaatgtctcacgaatttggaagtggccttaagtattattacaagaag	733
DB	815	ACTGGGCTGGCATCGGAGTCTCCAGAGTTGTCAGTCGGCACTGACCATTTATCGTCTTG	874
QY	734	tggcagattatattgctgcacacatttgaaaaagtgagggtgtccagtggaaaaatga	793
DB	875	TTGCCAATCATGTTGTAGTGATATCTCCCTAACCTGGAGGCTCTGTAGTCCAGAGAATAC	934
QY	794	gactaaagaaagacctgaaatctgagttctaaacagtgagattttggattggacatat	853
DB	935	GGTGGCCGATGAAGTGGAAAACCCGGGGATGAACAGTGGGATGCTTGGAGAAGACCTGA	994
QY	854	accaatactataaattttggcagaagaaggagatgttcagatacaagctctctctggac	913
DB	995	TTCAGTATTACCAGTTCCTAGCTGAGAAGGGTGAAGCTCCAAGCACAGGTTGGCCTGGAC	1054
QY	914	aattacatctaattggcagaaaggtctagatcaggatatactacaagcattacactact	973
DB	1055	AGCTGCATCTGCATGAGGGCTGGAGTAGAACAAANTCACACAGAGCGTTTGCCTACT	1114
QY	974	tcttaaggcagcaaaagccggggatgcaaatgcccattggcatttataggaaaagtatt	1033
DB	1115	TCAACTTAGCAGCAATGCTGGCAATTACATGCTATGGCCTTCTGGGAAGATGTATT	1174
QY	1034	tagaggggaatgctgcgctgcgcgaaaaataacgctactgccttccaagtactttccatgg	1093
DB	1175	CTGAAGGAAGTGACATCFGTACCTCAGAGTAATGAGACGGCACTTTCACACTTTTAAGAAG	1234
QY	1094	cagccgctaagggcaatgcaatcgcccttcattgggcttgcctcttacttctactcatgaa	1153
DB	1235	CTGCTGACATGGCAACCCCTGGGACAGAGCGGGCTTGAATGGCCTACCTCTACCGAA	1294
QY	1154	aaggagttcccctgaattatgcccraagcacttaataactttcagaagaagctgcgaaaaag	1213
DB	1295	GAGGCGTTCAAGTTATTTATGACCTGGCCCTCAAGTATTTCCAGAAAGCTGCTGAGCAAG	1354
QY	1214	ggggcccgacacacagctccagtttagcttcatgtactactctgacctggaatatgga	1273
DB	1355	GCTGGGTGACGGGCAGCTGCAGCTGGGCTCTATGTACTCAATGGCATTTGGATCAAGA	1414
QY	1274	aggattataaacttgccctcaaatatttttacctggcactccagatggggcagccccctcg	1333

Db	1415	GAGATTAATAAGCAGGCGTTGAAGATATTTTAATCTGGCTTCTCAAGGAGGCCATATCTTGG	14174
Qy	1334	ccattattatctggcgaagatgtatgcacagggaacaggagtagtaagatcatgcagaa	1393
Db	1475	CTTTCTATACCTTCGCACAGATGCACGCCAGCGGCACAGGGGTATGCGGTCTCTGTACA	1534
Qy	1394	ctgctgtggagctttataaaggtgtctgtgaactaggccaactgggctggaataatctctga	1453
Db	1535	CTGCAGTGGAGTTGTTTAAGATGTGTGTGACGAGGTGCGTGGTTCAGAGAGACTGATGA	1594
Qy	1454	cagcttacttggcctataagatggtgtgatatagattctctcttctgttcagtatgcactgc	1513
Db	1595	CTGCCTACAACAGCTATAAGGATGAGGACTACAACTGCTGCAGTGGTCCAGTACCTCTGC	1654
Qy	1514	tgcagaatlggggtatgaagtagctccaagcaatcagcattcaatttggaaatcctaaa	1573
Db	1655	TGGCTGAGCAGGCTACGAGGTGGCGCAGCAACGACGCTTTCATCTCCAGCCAGAGAG	1714
Qy	1574	aggctaaactcttgaagaagaagatgtatcccaatggcgtctctctatlggaatcgag	1633
Db	1715	AAGCAACCATTTGATGGTAGAATGAACCTTACCCAGAGCTTTACTGCAATGGAAACAGG	1774
Qy	1634	ctgccatcgaaggcaatgcatttgcctagagtataaattggagattaccattactatgctt	1693
Db	1775	CCGCCTCCCAAGGTTACACTGTGGCTAGAAATTTAAGCTTGGAGACTACCACCTTCTATGGCT	1834
Qy	1694	atgggactaagaagaactatcaaacagcagcacacactacagcattgtcagccaacaaat	1753
Db	1835	TTGGCACTGATGGGAATATGAGACGCGCATTTATTCATTACCGCGCTGGCTTCTCAGACGC	1894
Qy	1754	accacaacgcgcgaagcgaattgccaatctggcttatatgtatgaacaacggcttaggcata	1813
Db	1895	AGCACAGCGCCCAAGCTATGTTTAACCTGGGCTACATGCACGAGNAGGGCCCTAGGCATTA	1954
Qy	1814	caaaaggacattcacttggccagaagattgtacgacatggctgtcctcaaacgagtccagatg	1873
Db	1955	AACAGGACATTTCACTTGTCAAAACGGTTTTATGACATGGCAGCGGAAGCTAGCCAGATG	2014
Qy	1874	ccacatacctgtgctcttggcgtcatgaactgaaactgaaactacgcatttgcctcgagata	1933
Db	2015	CACAACTACCTGTGTTCCCTGGCACTCTGCAAAATTAGGTGTCGTATTTTCTTACAGTACA	2074
Qy	1934	tcctgttttttaatcagtttccaacgagatggaaactggctgaaactgtgaaactggcaacaccattg	1993
Db	2075	TACGGGAAGCAA--ACATTCGAGATCTATTACACACAACCTGGATATGGACCGCTTTTGG	2131
Qy	1994	gaccacactggggaactatttgtgattggcctcatg	2029
Db	2132	GACCCAGGTGGGACCTTTTACCTCATGACCATCATTTG	2167

RESULT	15
AX079694	
LOCUS	557 bp DNA
DEFINITION	Sequence 438 from Patent WO0107611.
ACCESSION	AX079694
VERSION	AX079694..1 GI:13159263
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1. (bases 1 to 557)
AUTHORS	Baker,K.P., Goddard,A. and Wood,W.I.
TITLE	Human polypeptides and methods for the use thereof
JOURNAL	Patent: WO 0107611-A 438 01-FEB-2001;
FEATURES	Genentech, Inc. (US) Location/Qualifiers 1..557 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	212 a 81 c 120 g 142 t
ORIGIN	2 others

Search completed: April 8, 2002, 21:31:00  
Job time: 3347·sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:42:09 ; Search time 121.25 Seconds

(without alignments)

14636.398 Million cell updates/sec

Title: US-09-714-882-1

Perfect score: 2070

Sequence: 1 atgaagccctgtctctgtt.....ttagaataccatgggttag 2070

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2069.6	100.0	2070	22 AAD06374	Novel human protei
2	2053.6	99.2	2067	22 AAD06375	Novel human protei
3	1650.2	79.7	1773	22 AAD06376	Novel human protei
4	1403.6	67.8	1500	22 AAD06378	Novel human protei
5	1403.6	67.8	1728	22 AAD06380	Novel human protei
6	1403.6	67.8	1731	22 AAD06379	Novel human protei
7	1254.6	60.6	1257	22 AAD06377	Novel human protei
8	633	30.6	7885	20 AAX76578	Human pancreas-spe
9	631.4	30.5	2382	20 AAX76577	Human pancreas-spe
10	631.4	30.5	8028	21 AAC75460	Human ORFX ORF1015
11	596.4	28.8	3666	20 AAX77013	Mouse Sel-1L splic

12	594.8	28.7	3667	20 AAX77012	Full length Mouse
13	510.4	24.7	557	22 AAF93617	cDNA encoding SRT
14	444	21.4	2144	22 AAF93859	Human cDNA encodin
15	437.6	21.1	2109	20 AAX77014	Human Sel-1L codin
16	316.4	15.3	1896	20 AAX77011	Partial Sel-1L cod
17	164	7.9	545	22 AAI16647	Probe #6580 for ge
18	164	7.9	545	22 AAI19859	Probe #545 used t
19	163	7.9	163	22 AAI25843	Probe #15776 for g
20	163	7.9	163	22 AAI52940	Probe #21626 used
21	149	7.2	936	22 AAF58252	Oligonucleotide D1
22	149	7.2	936	22 AAF58252	Oligonucleotide D1
23	149	7.2	936	22 AAF58254	Oligonucleotide D1
24	149	7.2	936	22 AAF58254	Oligonucleotide D1
25	149	7.2	936	22 AAF58257	Oligonucleotide D1
26	149	7.2	936	22 AAF58257	Oligonucleotide D1
27	149	7.2	936	22 AAF58259	Oligonucleotide D2
28	149	7.2	936	22 AAF58259	Oligonucleotide D2
29	149	7.2	936	22 AAF58262	Oligonucleotide D2
30	149	7.2	936	22 AAF58262	Oligonucleotide D2
31	149	7.2	938	22 AAF58255	Oligonucleotide D1
32	149	7.2	938	22 AAF58255	Oligonucleotide D1
33	146.4	7.1	1124	20 AAX61283	Human signal pept1
34	118.4	5.7	2225	21 AAZ52467	HTRM clone 1880692
35	117	5.7	1672	22 AAF92066	Human PRO1083 cDNA
36	117	5.7	1701	21 AAZ64982	Membrane-bound pro
37	117	5.7	1701	22 AAF44128	Human PRO4063 (UNQ
38	90.2	4.4	571	18 AAT85270	Expressed sequence
39	85.4	4.1	538	22 AAI17986	Probe #7919 for ge
40	85.4	4.1	538	22 AAI42976	Probe #11662 used
41	83.2	4.0	34001	22 AAF28531	Genomic fragment #
42	80.2	3.9	6644	20 AAX33181	Base sequence of t
43	80.2	3.9	7372	20 AAX33182	Base sequence of t
44	80.2	3.9	7797	20 AAX33180	Cowpox virus bsr f
45	80.2	3.9	7996	20 AAX33184	Base sequence of t

#### ALIGNMENTS

RESULT	1
AAD06374	
ID	AAD06374 standard; cDNA; 2070 BP.
XX	
AC	AAD06374;
XX	
DT	10-AUG-2001 (first entry)
XX	
DE	Novel human protein (NHP) cDNA #1, sharing similarity with Notch ligands.
XX	
KW	Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW	novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW	polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW	pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW	Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW	cholesterol metabolism; coronary artery disease; gene therapy;
KW	cerebroprotective; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2070
FT	/tag= a
FT	/product= "Novel human protein #1 (NHP)"
FT	1177
FT	/tag= b
FT	/note= "This ambiguity represents polymorphic site"
FT	1177..1179
FT	/tag= c
FT	note= "Encodes Glu"
XX	
XX	WO200136636-A2.
PN	
XX	
PD	25-MAY-2001.

XX 16-NOV-2000; 2000WO-US31373.  
PF  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX  
XX WPI: 2001-355635/37.  
DR P-PSDB; AAE02430.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
PS Claim 2; Page 26; 39pp; English.  
XX  
CC The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 2070 BP; 701 A; 379 C; 452 G; 537 T; 1 other;

Query Match 100.0%; Score 2069.6; DB 22; Length 2070;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2070; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY: 1 atgaagccctgtctctgtaataagagataatgataattcttgggtgcacaaataaaact 60  
DB 1 atgaagccctgtctctgtaataagagataatgataattcttgggtgcacaaataaaact 60

QY 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtcaccacaggtatca 120  
DB 61 atcaagcagaggaacataataaaagacaaaaggaagaaatgtcaccacaggtatca 120

QY 121 gtgaacgaatacaacaaatatttatcacacatatgtgaacaaagaacacatctagtaatga 180  
DB 121 gtgaacgaatacaacaaatatttatcacacatatgtgaacaaagaacacatctagtaatga 180

QY 181 atcaataaaagaaaatctcctggagaaaagaaagaaatcaacgtaaaaataagaataaaa 240  
DB 181 atcaataaaagaaaatctcctggagaaaagaaagaaatcaacgtaaaaataagaataaaa 240

QY 241 ggaattcaataaagatatcttgaagagaaaataagaatacatttacaagaagcaagcagag 300  
DB 241 ggaattcaataaagatatcttgaagagaaaataagaatacatttacaagaagcaagcagag 300

QY 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaaggtttctccagcag 360  
DB 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaaggtttctccagcag 360

QY 361 tctaaagcctaaacaaagaaagaaagcctacctacttttggcctaaagcagctgacatg 420  
DB 361 tctaaagcctaaacaaagaaagaaagcctacctacttttggcctaaagcagctgacatg 420

QY 421 ggaacctgaaagctatggagaaaaatggcgacgtcttctctatttggaaattttggcgtg 480  
DB 421 ggaacctgaaagctatggagaaaaatggcgacgtcttctctatttggaaattttggcgtg 480

QY 481 caaaatataacagcagctatccaattatatagctcttggctaaagaagatcatgtaa 540  
DB 481 caaaatataacagcagctatccaattatatagctcttggctaaagaagatcatgtaa 540

QY 541 gcccaaaacgcattagattttgtctcttcttgaatgaatgaatgaatgaatgaatgaat 600  
DB 541 gcccaaaacgcattagattttgtctcttcttgaatgaatgaatgaatgaatgaatgaat 600

QY 601 aaggcaactgatatttacacaccttggaaagtgtggaggaaacatgatgtcccagatgatt 660  
DB 601 aaggcaactgatatttacacaccttggaaagtgtggaggaaacatgatgtcccagatgatt 660

QY 661 ttgggttacagatatattgtcgggaatcaatgttctcacagaattgtgaagtgcctcaagt 720  
DB 661 ttgggttacagatatattgtcgggaatcaatgttctcacagaattgtgaagtgcctcaagt 720

QY 721 tattacaagaaagtgcagatttatattgtgacacatttgaacaaagtgaaggtgttcca 780  
DB 721 tattacaagaaagtgcagatttatattgtgacacatttgaacaaagtgaaggtgttcca 780

QY 781 gtggaaaaagtgcagtaacggaacacctgaaatcttgagttctaacagtgagattttg 840  
DB 781 gtggaaaaagtgcagtaacggaacacctgaaatcttgagttctaacagtgagattttg 840

QY 841 gattgggaacataaccataactataaattttggcgaagaaggagatgttcagatacaaa 900  
DB 841 gattgggaacataaccataactataaattttggcgaagaaggagatgttcagatacaaa 900

QY 901 gtctctcttggcaattacatctaatgtgcagaaaggcttagatcagattactacaaa 960  
DB 901 gtctctcttggcaattacatctaatgtgcagaaaggcttagatcagattactacaaa 960

QY 961 gattacacactcttcttaaggcagcaaaagccggagtgcaaatgccatggcattata 1020  
DB 961 gattacacactcttcttaaggcagcaaaagccggagtgcaaatgccatggcattata 1020

QY 1021 ggaagagatatttagaggggaatgctgcgtgcccgaataaacgctactgcttcaag 1080  
DB 1021 ggaagagatatttagaggggaatgctgcgtgcccgaataaacgctactgcttcaag 1080

QY 1081 tacttttccatggcagccagtaagggaatgcaatcgctcttcattgggttgcctctct 1140  
DB 1081 tacttttccatggcagccagtaagggaatgcaatcgctcttcattgggttgcctctct 1140

QY 1141 tactttcatggaagaggagttcccttgaaattatgcraagcacttaaatcttccagaaa 1200  
DB 1141 tactttcatggaagaggagttcccttgaaattatgcraagcacttaaatcttccagaaa 1200

QY 1201 gctgcgaaaaaggggtggcccagcacagttcccaattaggttcttactactctggc 1260  
DB 1201 gctgcgaaaaaggggtggcccagcacagttcccaattaggttcttactactctggc 1260

QY 1261 tctggaatatggaagattataaaactgcttcaataatatttaccctggcctctcagagt 1320  
DB 1261 tctggaatatggaagattataaaactgcttcaataatatttaccctggcctcagagt 1320

QY 1321 gggcagccccctgcgctatttattatcttgcgaagatgtatgcacaggaacagagtagta 1380  
DB 1321 gggcagccccctgcgctatttattatcttgcgaagatgtatgcacaggaacagagtagta 1380

QY 1381 agatcatgcagaactgctgtggagctttataaagggtgtctgtgaactagccaactggcct 1440  
DB 1381 agatcatgcagaactgctgtggagctttataaagggtgtctgtgaactagccaactggcct 1440

QY 1441 gagaaattctcagacagcttacttgcctataaggaatgggtgataatagattcttcttgtt 1500  
DB 1441 gagaaattctcagacagcttacttgcctataaggaatgggtgataatagattcttcttgtt 1500

QY 1501 cagtatgcactgcttgcaaaaaatggggtatgaagtgcctcaaaagcaattcagcattcatt 1560



Db 1501 cagtgacgcttcgcagaatgggcatgaagtgcataagcaattcagcattcatt 1560  
Qy 1561 ttggaattcaaaagcctaaacattcttgaaaaagagaagatgtatccaatggcgtcttc 1620  
Db 1561 ttggaattcaaaagcctaaacattcttgaaaaagagaagatgtatccaatggcgtcttc 1620  
Qy 1621 ctatggaatcagctgccattcaaggcaatgcatttgcctagtagtaaaatggagattac 1680  
Db 1621 ctatggaatcagctgccattcaaggcaatgcatttgcctagtagtaaaatggagattac 1680  
Qy 1681 cattactatgctatggactaaagaagactatcaaacagcagcagccacacactacagcatt 1740  
Db 1681 cattactatgctatggactaaagaagactatcaaacagcagcagccacacactacagcatt 1740  
Qy 1741 gcagccaaataaccacacgcgcgaagccatgttccaatctggcttatgtatgaacac 1800  
Db 1741 gcagccaaataaccacacgcgcgaagccatgttccaatctggcttatgtatgaacac 1800  
Qy 1801 ggcttaggcatacaaaagacattcaacttggccagaagattgtacgacatggctgctcaa 1860  
Db 1801 ggcttaggcatacaaaagacattcaacttggccagaagattgtacgacatggctgctcaa 1860  
Qy 1861 acgagtcagatgccacatacctgtgcttggcgttcgaactggaactggaactacgcat 1920  
Db 1861 acgagtcagatgccacatacctgtgcttggcgttcgaactggaactggaactacgcat 1920  
Qy 1921 ttgctccgggatactcgtttttaaatacagttccacacagagatgaaactgctgaaactg 1980  
Db 1921 ttgctccgggatactcgtttttaaatacagttccacacagagatgaaactgctgaaactg 1980  
Qy 1981 gacaacaccattggaccacactgggaactatttggattggcctcattgttctctgggctg 2040  
Db 1981 gacaacaccattggaccacactgggaactatttggattggcctcattgttctctgggctg 2040  
Qy 2041 attttgtgcttagaataccacatgggtag 2070  
Db 2041 attttgtgcttagaataccacatgggtag 2070

RESULT 2

AD06375  
ID AD06375 standard; cDNA; 2067 BP.

AC AAD06375;

DE 10-AUG-2001 (first entry)

XX Novel human protein (NHP) cDNA #2, sharing similarity with Notch ligands.

XX Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
FH CDS 1..2067

FT /\*tag= a  
FT /product= "Novel human protein #2 (NHP)"

FT 1177

FT misc\_feature /\*tag= b

FT /note= "This ambiguity represents polymorphic site"

FT 1177..1179

FT /\*tag= c

FT note= "Encodes Glu"

XX W0200136636-A2.

XX

PD 25-MAY-2001.

XX 16-NOV-2000; 2000WO-US31373.

XX 17-NOV-1999; 99US-0165959.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;

XX WPI; 2001-355635/37.

XX P-FSDB; AAE02431.

XX Novel isolated human polynucleotides encoding polypeptides sharing  
sequence similarity with mammalian SEL-1 proteins which are negative  
regulators of Notch family receptors, useful for treating diabetes,  
cancer

PS Claim 3; Page 28-29; 39pp; English.

XX The present sequence is a cDNA encoding novel human protein (NHP) which  
share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.

XX Sequence 2067 BP; 700 A; 378 C; 451 G; 537 T; 1 other;

Query Match 99.2%; Score 2053.6; DB 22; Length 2067;

Best Local Similarity 99.9%; Pred. NO. 0;

Matches 2067; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 atgaagcccttctctgttgaatagatattgataattcttgggtcacataaaact 60

Db 1 atgaagcccttctctgttgaatagatattgataattcttgggtcacataaaact 60

Qy 61 atcaagcagaggaacataataaagacaaaggaagaaatgtcaccacacaggtatca 120

Db 61 atcaagcagaggaacataataaagacaaaggaagaaatgtcaccacacaggtatca 120

Qy 121 gtgaacgaaatcaaacatatttatcacacatattgtggaacaaagacatctagtatga 180

Db 121 gtgaacgaaatcaaacatatttatcacacatattgtggaacaaagacatctagtatga 180

Qy 181 atcaataaagagaaatctcctggagaaagaaagaaatcaactgaaataaataa 240

Db 181 atcaataaagagaaatctcctggagaaagaaagaaatcaactgaaataaataa 240

Qy 241 ggaattcaaaataaagatatcttgaagagaataagaatcatttcaaaaagcagag 300

Db 241 ggaattcaaaataaagatatcttgaagagaataagaatcatttcaaaaagcagag 300

Qy 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaagggttctccagcag 360

Db 301 aaaaattttacagatgaaggagaccagctatttaagatgggcatcaagggttctccagcag 360

Qy 361 tctaaaagccaaaacaaagaaagaaagcctacctcttcttttggccaaagcagctgacatg 420

Db 361 tctaaaagccaaaacaaagaaagaaagcctacctcttcttttggccaaagcagctgacatg 420

```
QY 421 ggaaacttgaaagctatgagaaaaatggctgacgcttttctatttggaaattttggcgtg 480
DB 421 ggaaacttgaaagctatgagaaaaatggctgacgcttttctatttggaaattttggcgtg 480
QY 481 caaaatataacagcagctatccaattatatatagtgcttggctgcaaaaggagatcatgtaaa 540
DB 481 caaaatataacagcagctatccaattatatatagtgcttggctgcaaaaggagatcatgtaaa 540
QY 541 gccaaaacgcattaggattttgtctctctatgaaataggaaatggaatatatgaact 600
DB 541 gccaaaacgcattaggattttgtctctctatgaaataggaaatggaatatatgaact 600
QY 601 aaggcactgatattatcacacctttggaagtgtgaggaaacatgatgtcccagatgatt 660
DB 601 aaggcactgatattatcacacctttggaagtgtgaggaaacatgatgtcccagatgatt 660
QY 661 ttggggtacagatattttgcgggaatcaatgttctacagaattgtgaagttgcccctaagt 720
DB 661 ttggggtacagatattttgcgggaatcaatgttctacagaattgtgaagttgcccctaagt 720
QY 721 tattacaagaagtgagcagattatatctgacacatttgaataatgaaagtgtgtcca 780
DB 721 tattacaagaagtgagcagattatatctgacacatttgaataatgaaagtgtgtcca 780
QY 781 gtggaaaaagtggagtaacggaagacctgaaaatctgagttctaacaagtggagattttg 840
DB 781 gtggaaaaagtggagtaacggaagacctgaaaatctgagttctaacaagtggagattttg 840
QY 841 gattgggaacataaccaatactataaaattttggcagaagaaggagatgttcagatacaa 900
DB 841 gattgggaacataaccaatactataaaattttggcagaagaaggagatgttcagatacaa 900
QY 901 gtctctctggcaattacatactaatgtgcaggaaggtctagatcaggattactacaaa 960
DB 901 gtctctctggcaattacatactaatgtgcaggaaggtctagatcaggattactacaaa 960
QY 961 gcattacactactctttaaaggcagcaaggccgggagtgcaaatgccatggcatttata 1020
DB 961 gcattacactactctttaaaggcagcaaggccgggagtgcaaatgccatggcatttata 1020
QY 1021 ggaagatgtatttagaggggaatgctgccgtgcgcgcaaaaaataacgctactccctcaag 1080
DB 1021 ggaagatgtatttagaggggaatgctgccgtgcgcgcaaaaaataacgctactccctcaag 1080
QY 1081 tacttttccatggcagcagtaagggaatgcaatcgcccttcacgttggtcttctt 1140
DB 1081 tacttttccatggcagcagtaagggaatgcaatcgcccttcacgttggtcttctt 1140
QY 1141 tactttcatggaaaagggttcccttgaaattatgcraagcacttaatactttcagaaa 1200
DB 1141 tactttcatggaaaagggttcccttgaaattatgcraagcacttaatactttcagaaa 1200
QY 1201 gctgcggaataagggtggccgcagcaacagttccagttaggttcatgtactactctggc 1260
DB 1201 gctgcggaataagggtggccgcagcaacagttccagttaggttcatgtactactctggc 1260
QY 1261 tctggaatatggaagattatataacttgccttcaaatatttttacctggcactcagagt 1320
DB 1261 tctggaatatggaagattatataacttgccttcaaatatttttacctggcactcagagt 1320
QY 1321 gggcagccctgcctcatttattctatctggccaagatgtatgcacaggaacagagtagta 1380
DB 1321 gggcagccctgcctcatttattctatctggccaagatgtatgcacaggaacagagtagta 1380
QY 1381 agatcatgcagaactgctgtgagagtttataaagtgctgtgaaactaggccactgggct 1440
DB 1381 agatcatgcagaactgctgtgagagtttataaagtgctgtgaaactaggccactgggct 1440
QY 1441 gagaattctctgacagctactttgcctataagagatgtatagattcttctcttatt 1500
DB 1441 gagaattctctgacagctactttgcctataagagatgtatagattcttctcttatt 1500
```

```
QY 1501 cagtatgcactgcttcagaaaaatggggtatgaagttagctcaaaagcaattcagcattcatt 1560
DB 1501 cagtatgcactgcttcagaaaaatggggtatgaagttagctcaaaagcaattcagcattcatt 1560
QY 1561 ttggaatctaaaaaggctaaacatttcttgaataaaggagaatgtatccaatggcgtctctc 1620
DB 1561 ttggaatctaaaaaggctaaacatttcttgaataaaggagaatgtatccaatggcgtctctc 1620
QY 1621 ctatggaatcgagctggccattcaaggcaatgcatcttctagagtagtaaaatttgagagattac 1680
DB 1621 ctatggaatcgagctggccattcaaggcaatgcatcttctagagtagtaaaatttgagagattac 1680
QY 1681 cattactatggctatgggactaagaagaactatcaaacagcagccacacactacagcatt 1740
DB 1681 cattactatggctatgggactaagaagaactatcaaacagcagccacacactacagcatt 1740
QY 1741 gcagccaacaataaccacaacgcgaagccatgttcaatcttggcttatgtatgaacac 1800
DB 1741 gcagccaacaataaccacaacgcgaagccatgttcaatcttggcttatgtatgaacac 1800
QY 1801 ggcttaggcatacaaaaggacattcacttggccagaagattgtacgacatggctgctcaa 1860
DB 1801 ggcttaggcatacaaaaggacattcacttggccagaagattgtacgacatggctgctcaa 1860
QY 1861 acgagtccagatgcccacatacctgtcttcttgcgtcatgaaactggaactacgcatt 1920
DB 1861 acgagtccagatgcccacatacctgtcttcttgcgtcatgaaactggaactacgcatt 1920
QY 1921 ttgctccgggatactcgttttttaatacagttcaaacagagatggaactggctgaaactg 1980
DB 1921 ttgctccgggatactcgttttttaatacagttcaaacagagatggaactggctgaaactg 1980
QY 1981 gacaacacattggccacactggacattattgttgatggcctcattgttctctggcgtg 2040
DB 1981 gacaacacattggccacactggacattattgttgatggcctcattgttctctggcgtg 2040
QY 2041 attttgttcttagaataaccattgggtag 2070
DB 2038 attttgttcttagaataaccattgggtag 2067

RESULT 3
AAD06376
ID AAD06376 standard; cDNA; 1773 BP.
XX AC AAD06376;
XX DF 10-AUG-2001 (first entry)
XX DE Novel human protein (NHP) cDNA #3, sharing similarity with Notch ligands.
KW Human; nootropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;
KW cholesterol metabolism; coronary artery disease; gene therapy;
KW cerebroprotective; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..1773
XX FT /*tag= a
XX FT /product= "Novel human protein #3 (NHP)"
XX FT misc_feature 1177
XX FT /*tag= b
XX FT /note= "This ambiguity represents polymorphic site"
XX FT 1177..1179
XX FT /*tag= c
XX FT note= "Encodes Glu"
XX PN WO200136636-A2.
```



QY 1501 cagtgcactgctgcagaaatgggtatgaagtgcctcaagcaattcagcattcatt 1560  
 |||||  
 Db 1501 cagtgcactgctgcagaaatgggtatgaagtgcctcaagcaattcagcattcatt 1560  
 |||||  
 QY 1561 ttggaatctaaagcgaatactcttgaagaaagagaagatgtatccaatgctcttc 1620  
 |||||  
 Db 1561 ttggaatctaaagcgaatactcttgaagaaagagaagatgtatccaatgctcttc 1620  
 |||||  
 QY 1621 ctatggaatcgagctgcattcaagcgaatgctttg 1657  
 |||||  
 Db 1621 ctatggaatcgagctgcattcaagcgaatgctttg 1657  
 |||||

RESULT 4  
 AAD06378  
 ID AAD06378 standard; cDNA; 1500 BP.  
 XX  
 AC AAD06378;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Novel human protein (NHP) cDNA #5, sharing similarity with Notch ligands.  
 XX  
 KW Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
 KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
 KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
 KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
 KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
 KW cholesterol metabolism; coronary artery disease; gene therapy;  
 KW cerebroprotective; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..1500  
 FT /\*tag= a  
 FT /product= "Novel human protein #5 (NHP)"  
 FT 1177  
 FT misc\_feature /\*tag= b  
 FT /note= "This ambiguity represents polymorphic site"  
 FT 1177..1179  
 FT /\*tag= c  
 FT note= "Encodes Glu"  
 FT  
 XX WO200136636-A2.  
 XX  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 16-NOV-2000; 2000WO-US31373.  
 XX  
 PR 17-NOV-1999; 99US-0165959.  
 XX  
 PA (LEXI-) LEXICON GENETICS INC.  
 XX  
 PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
 XX  
 DR WPI; 2001-355635/37.  
 DR P-PSDB; AAE02434.  
 XX  
 PT Novel isolated human polynucleotides encoding polypeptides sharing  
 PT sequence similarity with mammalian SEL-1 proteins which are negative  
 PT regulators of Notch family receptors, useful for treating diabetes,  
 PT cancer  
 XX  
 PS Claim 6; Page 33-34; 39pp; English.  
 XX  
 CC The present sequence is a cDNA encoding novel human protein (NHP) which  
 CC share structural similarity with animal Notch ligands, particularly  
 CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
 CC Notch receptors and their associated signalling pathways have been  
 CC associated with development, apoptosis, neuron growth and maintenance.  
 CC Labeled NHP probes can be used to screen a human genomic library which

CC is helpful for identifying polymorphisms, determining the genomic  
 CC structure of a given locus/allele and designing diagnostic tests. The  
 CC NHP is also useful in screening techniques for drugs which treats  
 CC symptomatic or phenotypic manifestations of perturbing the normal  
 CC function of NHP in the body. Nucleotide constructs encoding functional  
 CC NHPs, antisense, antisense molecules can be used in gene therapy  
 CC approaches for modulating gene expression such as for preventing or  
 CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
 CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
 CC such as Parkinson's disease, stroke, vascular dementia and conditions  
 CC requiring modulation of fat and cholesterol metabolism such as coronary  
 CC artery disease.  
 XX  
 SQ Sequence 1500 BP; 537 A; 254 C; 328 G; 380 T; 1 other;

Query Match 67.8%; Score 1403.6; DB 22; Length 1500;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagcccttctcttgaatagagatattgataattcttgggtcacattaaaact 60  
 |||||  
 Db 1 atgaagcccttctcttgaatagagatattgataattcttgggtcacattaaaact 60  
 |||||  
 QY 61 atcaagcagaggaacataataaaagacaaaagaaagaaatgtcaccacacaggtatca 120  
 |||||  
 Db 61 atcaagcagaggaacataataaaagacaaaagaaagaaatgtcaccacacaggtatca 120  
 |||||  
 QY 121 gtgaacgaatacaacaataattttcacacatttggaacaaagaacatctgtagta 180  
 |||||  
 Db 121 gtgaacgaatacaacaataattttcacacatttggaacaaagaacatctgtagta 180  
 |||||  
 QY 181 atcaataaaagaaaatctctgagaaaagaaaagaaatcaacgtataaataaaa 240  
 |||||  
 Db 181 atcaataaaagaaaatctctgagaaaagaaaagaaatcaacgtataaataaaa 240  
 |||||  
 QY 241 ggaattcaaaataaagatatcttgaagagaataagaatacatttacaagaacagcagag 300  
 |||||  
 Db 241 ggaattcaaaataaagatatcttgaagagaataagaatacatttacaagaacagcagag 300  
 |||||  
 QY 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaaggttctccagcag 360  
 |||||  
 Db 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaaggttctccagcag 360  
 |||||  
 QY 361 tctaaaagccaaacaaacaaagaaagaaagcctaccttcttgcacaaagcagctgacatg 420  
 |||||  
 Db 361 tctaaaagccaaacaaacaaagaaagaaagcctaccttcttgcacaaagcagctgacatg 420  
 |||||  
 QY 421 ggaaacttgaaagctatggagaaaatggctgcgctttgtctatttggaaatttggcgctg 480  
 |||||  
 Db 421 ggaaacttgaaagctatggagaaaatggctgcgctttgtctatttggaaatttggcgctg 480  
 |||||  
 QY 481 caaataataacagcagctatccaaattatagctcttggcttgaagaagagatcatgtaa 540  
 |||||  
 Db 481 caaataataacagcagctatccaaattatagctcttggcttgaagaagagatcatgtaa 540  
 |||||  
 QY 541 gcccaaacgcattaggattttgtctcttctgaataggaatggaatgatcatcaagct 600  
 |||||  
 Db 541 gcccaaacgcattaggattttgtctcttctgaataggaatggaatgatcatcaagct 600  
 |||||  
 QY 601 aaggcactgatattacacctttggaagtgtcggagaaacatgatgtcccagattgatt 660  
 |||||  
 Db 601 aaggcactgatattacacctttggaagtgtcggagaaacatgatgtcccagattgatt 660  
 |||||  
 QY 661 ttggggtacagatatgttcgggaataatgttctacagaattgtgaagtgtccctaaagt 720  
 |||||  
 Db 661 ttggggtacagatatgttcgggaataatgttctacagaattgtgaagtgtccctaaagt 720  
 |||||  
 QY 721 tattacaagaagtgcagattatttctgcacacatttgaaaaaagtgaaagtggttcca 780  
 |||||  
 Db 721 tattacaagaagtgcagattatttctgcacacatttgaaaaaagtgaaagtggttcca 780  
 |||||  
 QY 781 gtggaaaaagtgaagactaacggaagacctgaaaaactctgagttctaaacagtgagatttg 840

Db 781 |gtggaaaaagtgagactaacgaaagacctgaaatctgagttctaaacagtgagatttg 840  
QY 841 |gattggacataccaatactataaatttttggcagaagaagagagattgtccagatacaa 900  
Db 841 |gattggacataccaatactataaatttttggcagaagaagagagattgtccagatacaa 900  
QY 901 |gtctctcttggacaattacatactaatgtgagcagaagaaggtctagatcaggattactacaaa 960  
Db 901 |gtctctcttggacaattacatactaatgtgagcagaagaaggtctagatcaggattactacaaa 960  
QY 961 |gcattacactacttcttaaaagcagaagccggcgagggtgcaaatgcaatgccttctcaag 1020  
Db 961 |gcattacactacttcttaaaagcagaagccggcgagggtgcaaatgcaatgccttctcaag 1020  
QY 1021 |gaaagagatgatttagaggggaatgctgcgtgccgcgaataaataacgcctactgcttcaag 1080  
Db 1021 |gaaagagatgatttagaggggaatgctgcgtgccgcgaataaataacgcctactgcttcaag 1080  
QY 1081 |tactttccatggcagccagtaaggcgaatgcaatgcgccttcattggtggtcttctt 1140  
Db 1081 |tactttccatggcagccagtaaggcgaatgcaatgcgccttcattggtggtcttctt 1140  
QY 1141 |tactttcattgaaaaggaggtcccttggaattatgcraagcaacttaatactttcagaaa 1200  
Db 1141 |tactttcattgaaaaggaggtcccttggaattatgcraagcaacttaatactttcagaaa 1200  
QY 1201 |gctgcgaaaaaggtggtccgcagcacagctccagttcaggttaggtctcatgtaactactctggc 1260  
Db 1201 |gctgcgaaaaaggtggtccgcagcacagctccagttcaggttaggtctcatgtaactactctggc 1260  
QY 1261 |tgtgaatatggaagattataaaatgcttcctcaaatatttttaactggcattcagagt 1320  
Db 1261 |tgtgaatatggaagattataaaatgcttcctcaaatatttttaactggcattcagagt 1320  
QY 1321 |ggcgagccctgcgacttatttatctgccaagatgtatgcaacaggaacagagtagta 1380  
Db 1321 |ggcgagccctgcgacttatttatctgccaagatgtatgcaacaggaacagagtagta 1380  
QY 1381 |agatcatgcagaactgctgtggag 1404  
Db 1381 |agatcatgcagaactgctgtggag 1404

RESULT 5  
AAD06380  
ID AAD06380 standard; cDNA; 1728 BP.  
XX  
AC AAD06380;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE Novel human protein (NHP) cDNA #7, sharing similarity with Notch ligands.  
XX  
KW Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;  
KW novel human protein; Notch ligand; apoptosis; neuron growth; therapy;  
KW polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;  
KW pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;  
KW Parkinson's disease; stroke; vascular dementia; fat metabolism;  
KW cholesterol metabolism; coronary artery disease; gene therapy;  
KW cerebroprotective; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1728  
FT /\*tag= a  
FT /product= "Novel human protein #7 (NHP)"  
FT misc\_feature 1177  
FT /\*tag= b  
FT /note= "This ambiguity represents polymorphic site"  
FT unsure 1177..1179  
FT /\*tag= c

FT note= "Encodes Glu"  
PN WO200136636-A2.  
XX  
PD 25-MAY-2001.  
XX  
PF 16-NOV-2000; 2000WO-US31373.  
XX  
PR 17-NOV-1999; 99US-0165959.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
PI Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;  
XX  
DR WPI; 2001-355635/37.  
XX  
DR P-PSDB; AAE02436.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer -  
XX  
PS Disclosure; Page 37; 39pp; English.  
XX  
CC The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treat  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy or  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 1728 BP; 592 A; 309 C; 380 G; 446 T; 1 other;  
  
Query Match 67.8%; Score 1403.6; DB 22; Length 1728;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 atgaagccctgtctctgttaatagagatattgataattcttgggtgcacaattaaaact 60  
Db 1 atgaagccctgtctctgttaatagagatattgataattcttgggtgcacaattaaaact 60  
  
QY 61 atcaagcagaggaacataataaaagacaaaaggaagaataatgtcaccacacaggtatca 120  
Db 61 atcaagcagaggaacataataaaagacaaaaggaagaataatgtcaccacacaggtatca 120  
  
QY 121 gtgaacgaatacaacaattatttaccacacatattggaacaaacacatctagtattga 180  
Db 121 gtgaacgaatacaacaattatttaccacacatattggaacaaacacatctagtattga 180  
  
QY 181 atcaataaaagagaaaatctcctggagaaaagaagaatacaacgtaaaataagaataaaa 240  
Db 181 atcaataaaagagaaaatctcctggagaaaagaagaatacaacgtaaaataagaataaaa 240  
  
QY 241 ggaattcaaaataaagatatcttgaagagagaataagaatacatttacaacaaagcagag 300  
Db 241 ggaattcaaaataaagatatcttgaagagagaataagaatacatttacaacaaagcagag 300  
  
QY 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaagggttctccagcag 360  
Db 301 aaaaattttacagatgaagagaccagctatttaagatgggcatcaagggttctccagcag 360

Qy	361	tctaaagccaaaaacaaaaagaagagcctacactcttcttgccaaagcagctgacatg	420
Db	361	tctaaagccaaaaacaaaaagaagagcctacactcttcttgccaaagcagctgacatg	420
Qy	421	ggaaactgaaagctatgagaaaaatgctaacgctttgctatttgaaaatttggcgtg	480
Db	421	ggaaactgaaagctatgagaaaaatgctaacgctttgctatttgaaaatttggcgtg	480
Qy	481	caaaataaacagcagctatccaattatgatgctcttggctaaagaagatcatgtaaa	540
Db	481	caaaataaacagcagctatccaattatgatgctcttggctaaagaagatcatgtaaa	540
Qy	541	gccccaaagcattagagattttgtctctcttatggaataggaaatggaaatgatcaagct	600
Db	541	gccccaaagcattagagattttgtctctcttatggaataggaaatggaaatgatcaagct	600
Qy	601	aaggcactgatatactacaccttggaaagtgcctgaggaacaatgatgtcccagatgatt	660
Db	601	aaggcactgatatactacaccttggaaagtgcctgaggaacaatgatgtcccagatgatt	660
Qy	661	ttggggtacagatatttgtcgggaatcaattgtctacagaattgtgaagttgcctaaagt	720
Db	661	ttggggtacagatatttgtcgggaatcaattgtctacagaattgtgaagttgcctaaagt	720
Qy	721	tattacaagaagtgcagattatattgtctgacacatttgaaaaagtgaagtggttcca	780
Db	721	tattacaagaagtgcagattatattgtctgacacatttgaaaaagtgaagtggttcca	780
Qy	781	gtggaaaaagtgcagctaacgcgaagacactgaaaatctgagttcttaacagtgagatttg	840
Db	781	gtggaaaaagtgcagctaacgcgaagacactgaaaatctgagttcttaacagtgagatttg	840
Qy	841	gattgggacatataccaatactataaaatttttggcagaaaagagagatgttcagatacaa	900
Db	841	gattgggacatataccaatactataaaatttttggcagaaaagagagatgttcagatacaa	900
Qy	901	gtctctcttgacaattacatctaatttggcgagaaagtctagatcaggattactacaaa	960
Db	901	gtctctcttgacaattacatctaatttggcgagaaagtctagatcaggattactacaaa	960
Qy	961	gcattacactactctttaaagcgacaaagccgggagtgcaaatgccatggcatttata	1020
Db	961	gcattacactactctttaaagcgacaaagccgggagtgcaaatgccatggcatttata	1020
Qy	1021	ggaaaagtctatttagaggggaatgctgcgcgtgcgcgcaaaataacgctactgcctcaag	1080
Db	1021	ggaaaagtctatttagaggggaatgctgcgcgtgcgcgcaaaataacgctactgcctcaag	1080
Qy	1081	tacttttccatgycagccagtaaggcgcaatgcaatcgcccttcattggcttgctctctt	1140
Db	1081	tacttttccatgycagccagtaaggcgcaatgcaatcgcccttcattggcttgctctctt	1140
Qy	1141	tactttcatgaaaaaggaattcccttgaattatgccraagcaccttaaatcttcagaaa	1200
Db	1141	tactttcatgaaaaaggaattcccttgaattatgccraagcaccttaaatcttcagaaa	1200
Qy	1201	gctgcgaaaaaagggtggcccgacgcacagttccagttaggcttcatgtactactctggc	1260
Db	1201	gcctgcgaaaaaagggtggcccgacgcacagttccagttaggcttcatgtactactctggc	1260
Qy	1261	tcttggaaatggaagattataaaattgccttcaaatatttttaccctggcatctcagagt	1320
Db	1261	tcttggaaatggaagattataaaattgccttcaaatatttttaccctggcatctcagagt	1320
Qy	1321	ggcgagccctctgcactttatcatctggcccaagatgtatgcaacaggaacagagtagta	1380
Db	1321	ggcgagccctctgcactttatcatctggcccaagatgtatgcaacaggaacagagtagta	1380
Qy	1381	agatcatgcgaactgctgtggag	1404
Db	1381	agatcatgcgaactgctgtggag	1404

RESULT	6	
AAD06379		
ID	AAD06379	standard; cDNA; 1731 BP.
XX		
AC	AAD06379;	
XX		
DT	10-AUG-2001	(first entry)
XX		
DE	Novel human protein (NHP) cDNA #6, sharing similarity with Notch ligands.	
XX		
KW	Human; nontropic; neuroprotective; antidiabetic; cytostatic; SEL-1; NHP;	
KW	novel human protein; Notch ligand; apoptosis; neuron growth; therapy;	
KW	polymorphism; drug screening; Alzheimer's disease; diabetes; cancer;	
KW	pancreatic cancer; insulinoma; blood pressure; neurodegenerative disease;	
KW	Parkinson's disease; stroke; vascular dementia; fat metabolism;	
KW	cholesterol metabolism; coronary artery disease; gene therapy;	
KW	cerebroprotective; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1731
FT		/*tag= a
FT		/product= "Novel human protein #6 (NHP)"
FT	misc_feature	1177
FT		/*tag= b
FT		/note= "This ambiguity represents polymorphic site"
FT	unsure	1177..1179
FT		/*tag= C
FT		note= "Encodes Glu"
XX		
PN	WO200136636-A2.	
XX		
PD	25-MAY-2001.	
XX		
PF	16-NOV-2000; 2000WO-US31373.	
XX		
PR	17-NOV-1999; 99US-0165959.	
XX		
PA	(LEXI-) LEXICON GENETICS INC.	
XX		
PI	Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;	
XX		
XX	WPI; 2001-355635/37.	
DR	P-PSDB; AAE02435.	
XX		
PT	Novel isolated human polynucleotides encoding polypeptides sharing	
PT	sequence similarity with mammalian SEL-1 proteins which are negative	
PT	regulators of Notch family receptors, useful for treating diabetes,	
PT	cancer	
XX		
PS	Disclosure; Page 35-36; 39pp; English.	
XX		
CC	The present sequence is a cDNA encoding novel human protein (NHP) which	
CC	share structural similarity with animal Notch ligands, particularly	
CC	SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.	
CC	Notch receptors and their associated signalling pathways have been	
CC	associated with development, apoptosis, neuron growth and maintenance.	
CC	Labeled NHP probes can be used to screen a human genomic library which	
CC	is helpful for identifying polymorphisms, determining the genomic	
CC	structure of a given locus/allele and designing diagnostic tests. The	
CC	NHP is also useful in screening techniques for drugs which treats	
CC	symptomatic or phenotypic manifestations of perturbing the normal	
CC	function of NHP in the body. Nucleotide constructs encoding functional	
CC	NHPs, antisense, antisense molecules can be used in gene therapy	
CC	approaches for modulating gene expression such as for preventing or	
CC	treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,	
CC	insulinomas), blood pressure abnormalities, neurodegenerative diseases	
CC	such as Parkinson's disease, stroke, vascular dementia and conditions	
CC	requiring modulation of fat and cholesterol metabolism such as coronary	
CC	artery disease.	
XX		



XX WPI: 2001-355635/37.  
DR P-PSDB; AAE02433.  
XX  
PT Novel isolated human polynucleotides encoding polypeptides sharing  
PT sequence similarity with mammalian SEL-1 proteins which are negative  
PT regulators of Notch family receptors, useful for treating diabetes,  
PT cancer  
XX  
PS Claim 5; Page 32; 39pp; English.  
XX  
CC The present sequence is a cDNA encoding novel human protein (NHP) which  
CC share structural similarity with animal Notch ligands, particularly  
CC SEL-1. SEL-1 proteins are negative regulators of Notch family receptors.  
CC Notch receptors and their associated signalling pathways have been  
CC associated with development, apoptosis, neuron growth and maintenance.  
CC Labeled NHP probes can be used to screen a human genomic library which  
CC is helpful for identifying polymorphisms, determining the genomic  
CC structure of a given locus/allele and designing diagnostic tests. The  
CC NHP is also useful in screening techniques for drugs which treats  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of NHP in the body. Nucleotide constructs encoding functional  
CC NHPs, antisense, antisense molecules can be used in gene therapy  
CC approaches for modulating gene expression such as for preventing or  
CC treating Alzheimer's disease, diabetes, cancer (pancreatic cancer,  
CC insulinomas), blood pressure abnormalities, neurodegenerative diseases  
CC such as Parkinson's disease, stroke, vascular dementia and conditions  
CC requiring modulation of fat and cholesterol metabolism such as coronary  
CC artery disease.  
XX  
SQ Sequence 1257 BP; 464 A; 206 C; 270 G; 316 T; 1 other;

Query Match 60.6%; Score 1254.6; DB 22; Length 1257;  
Best Local Similarity 100.08; Pred. No. 9.9e-292;  
Matches 1257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagccctgtctctgttaagagatattgataattcttgggtgcacaaattaaaaact 60  
DB 1 atgaagccctgtctctgttaagagatattgataattcttgggtgcacaaattaaaaact 60

QY 61 atcaagcagaggaacataataaaagacaaagaaagaaatgtccacacacagtatca 120  
DB 61 atcaagcagaggaacataataaaagacaaagaaagaaatgtccacacacagtatca 120

QY 121 ggaacgaatacaacaattattacacacatttggaacaaagaaacatctagtaatga 180  
DB 121 ggaacgaatacaacaattattacacacatttggaacaaagaaacatctagtaatga 180

QY 181 atcaataaaagaaatactctctgagaaaaagaaagaaatcaacgtataaataaagaa 240  
DB 181 atcaataaaagaaatactctctgagaaaaagaaagaaatcaacgtataaataaagaa 240

QY 241 ggaattcaaaataaagatattcttgagagaaataagaatcatttcaaaaagcagcag 300  
DB 241 ggaattcaaaataaagatattcttgagagaaataagaatcatttcaaaaagcagcag 300

QY 301 aaaaattttacagatgaagagacagcattttaaagatgggcatcaaggttctccagcag 360  
DB 301 aaaaattttacagatgaagagacagcattttaaagatgggcatcaaggttctccagcag 360

QY 361 tctaaaagccaaaaacaaagaaagcctacctacttttggcaaaagcagctgacatg 420  
DB 361 tctaaaagccaaaaacaaagaaagcctacctacttttggcaaaagcagctgacatg 420

QY 421 ggaacttgaaagctatggagaaatgctgacgcttggctatttggaaaatttggcggtg 480  
DB 421 ggaacttgaaagctatggagaaatgctgacgcttggctatttggaaaatttggcggtg 480

QY 481 caaataaagcagcctatcaattatatagcttggctaaagaggaatcatgtaaa 540  
DB 481 caaataaagcagcctatcaattatatagcttggctaaagaggaatcatgtaaa 540

QY 541 gcccaaacgattaggtttttgtcttcttatggaataggaaatgatcaagct 600  
DB 541 gcccaaacgattaggtttttgtcttcttatggaataggaaatgatcaagct 600

QY 601 aagcactgatatattacaccttttggaagtgctgagagaaacatgatgtcccagatgatt 660  
DB 601 aagcactgatatattacaccttttggaagtgctgagagaaacatgatgtcccagatgatt 660

QY 661 ttgggttacagatatattgtcgaggaaatcaatgtttctacagaatttgaaagttgccctaagt 720  
DB 661 ttgggttacagatatattgtcgaggaaatcaatgtttctacagaatttgaaagttgccctaagt 720

QY 721 tattacaagaagtgcagatttatattgtcacacatttgaaacattgaaagtgaaagtttcca 780  
DB 721 tattacaagaagtgcagatttatattgtcacacatttgaaacattgaaagtgaaagtttcca 780

QY 781 gttgaaaaagtgcagactaacgaaagcctgaaatctgagttctaaacagtgagatttg 840  
DB 781 gttgaaaaagtgcagactaacgaaagcctgaaatctgagttctaaacagtgagatttg 840

QY 841 gattggacatatataactataataatttttgcagaaagagagagattgttcagatacaa 900  
DB 841 gattggacatatataactataataatttttgcagaaagagagagattgttcagatacaa 900

QY 901 gtctctcttgacaaattacataatttgagcagaaaggtcttagatcaggattactacaaa 960  
DB 901 gtctctcttgacaaattacataatttgagcagaaaggtcttagatcaggattactacaaa 960

QY 961 gattacactacttcttaaaagcagaaagcggggagtgcaaatgccatggcatttata 1020  
DB 961 gattacactacttcttaaaagcagaaagcggggagtgcaaatgccatggcatttata 1020

QY 1021 ggaagagattatttagagggaatgctccgtgcccgaataaactactgctcttcaag 1080  
DB 1021 ggaagagattatttagagggaatgctccgtgcccgaataaactactgctcttcaag 1080

QY 1081 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggttcttctt 1140  
DB 1081 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggttcttctt 1140

QY 1141 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggttcttctt 1200  
DB 1141 tacttttccatggcagccagtaagggcaatgcaatcgcccttcattgggttcttctt 1200

QY 1201 gctgcgaaaaaggggtggccgacgcacagttccagtttaggcttcatgtactact 1255  
DB 1201 gctgcgaaaaaggggtggccgacgcacagttccagtttaggcttcatgtactact 1255

RESULT 8  
AA76578  
ID AA76578 standard; cDNA; 7885 BP.  
XX AA76578;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human pancreas-specific tumour suppressor TSA305 gene SEQ ID NO:3.  
XX  
KW Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;  
KW therapy; tumour suppressor; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 46..2430  
FT /\*tag= a  
XX  
PN WO9928457-A1.  
XX  
PD 10-JUN-1999.  
XX  
PF 25-NOV-1998; 98WO-JP05306.



XX 20-APR-1998; 98JP-0126803.  
 PR 28-NOV-1997; 97JP-0343789.  
 XX (SAKA ) OTSUKA PHARM CO LTD.  
 PA Harada Y, Ozaki K;  
 XX WPI: 1999-358128/30.  
 XX P-PSDB: AAY17750.  
 DR Tumour suppressor gene TSA305 expressed specifically in pancreas,  
 PT useful for diagnosis and treatment of pancreatic cancer  
 XX Disclosure; Page 43-52; 54pp; Japanese.  
 XX The present sequence represents a gene designated TSA305, which is,  
 CC expressed specifically in pancreas tissue. The gene, sequences  
 CC hybridizing with it, its expression product, and antibodies recognizing  
 CC the expression product, are useful in the investigation, diagnosis,  
 CC prevention and treatment of pancreatic cancer.  
 XX Sequence 7885 BP; 2332 A; 1466 C; 1619 G; 2468 T; 0 other;  
 SQ

Query Match 30.6%; Score 633; DB 20; Length 7885;  
 Best Local Similarity 60.6%; Pred. No. 3.3e-142;  
 Matches 1057; Conservative 0; Mismatches 685; Indels 3; Gaps 1;

Qy 285 acaaaagcaagcagagaaattttacagatgaaggagaccagctatttaagatgggcat 344  
 Db 558 agaagagcgtgaagagacgcagatgcaggaagcagagaaatgatgtatcaaaactggaat 617  
 Qy 345 caaggttcagcagctaaagcccaaaacaaagaaagaaagcctaccctattttgyc 404  
 Db 618 gaaatccttaatggaagcaataagaagccaaagaaagaaagacatcgtgtatctcca 677  
 Qy 405 caaagcagctgacatgggaactgaaagctatggagaaatggctgcgctgtgctatt 464  
 Db 678 aaagcagcagcagatgaaccatacaaaagccctggagagtgctatgctctttatt 737  
 Qy 465 tggaaattttggcgtgcaaaatataaacagcagctatcccaattatagtgcttggctaa 524  
 Db 738 tgggtattacttgccacgaatataccagcagcagagagatgtttgagagctgactga 797  
 Qy 525 agaagatcatgtaaaagcccaaacgcattagagattttgtcttcttatgaataaggaat 584  
 Db 798 ggaagcctctcccaaggagcagactgctcttggccttctgtatgctctggacttggctg 857  
 Qy 585 ggaatgatcaagcgaagcagctgatatattacacctttggaagtgctggaggaacat 644  
 Db 858 taattcaagtcaggcaaaagccttgatattatatacatattggagctcttggggcaatct 917  
 Qy 645 gatgtccagatgattttggggtacagatatttgcgggaatcaatgtctacagaattg 704  
 Db 918 aatagccacatggtttgggttacagatactggcgtggcctgcctccagagttg 977  
 Qy 705 tgaagttgcccataattacagaagaagtggcagattatattgctgacacattgaaaa 764  
 Db 978 tgaatcctccctgactcactatcgctctgttggccaatcatgtctgtatgtatctcgt 1037  
 Qy 765 aagtgaaagtggttccagtggaagagtgagactaaacggaagacacctgaaatctgagttc 824  
 Db 1038 aacaggaggtcagtagtacagagaatacggctgctgctgatgaagggaataccaggaat 1097  
 Qy 825 taacagtgagattttggattgggacataataccaataactataaatttttggcagaagaagg 884  
 Db 1098 gaacagtggaatgctagaagaagatttgattcoattatcaccagtttccctagctgaaaaagg 1157  
 Qy 885 agatgttcagatacaagctctctctggacaattatcatataattggcagaagaaggtctaga 944  
 Db 1158 tgaatcaaacgacaggttggctcttggcaaacctgcaactgcaagggagggcgtggagtaga 1217

Qy 945 tcagattactacaagcattacactacttcttaaaagcagcaaaagccggagtgcaaa 1004  
 Db 1218 acagaatcatcagagcatttgactcaatttagcagcaaatgtggaattcaca 1277  
 Qy 1005 tgcattggcatttatagaaaagatgtatttagaggggaatgctccgtgcgcgaataaa 1064  
 Db 1278 tgcattggcattttggaaagatgtatttagaggggaatgacattgtacctcagagtaa 1337  
 Qy 1065 cgtactgctctcaagtacttttccatggcagccagtaagggcgaatgcaatcggccttca 1124  
 Db 1338 tgagacagctctccactactttaagaaagctgctgacatggcgaaccagttgacagag 1397  
 Qy 1125 tggccttggtctcttacttctatgaaaaggagttccctcgaattatgcccraagcaat 1184  
 Db 1398 tggccttggaatgacctctatgggagagaggttcaagttaattatgctcagccct 1457  
 Qy 1185 taatactttcagaaagctgcggaaggggtggccgcagcacacagttccagttaggctt 1244  
 Db 1458 taagtattccagaagagctgctgaacaggctgggtgagtgagcgtcacagcttgggttc 1517  
 Qy 1245 catgtactactctgctctggaatggaagattataaaacttgcttcaaatatttta 1304  
 Db 1518 catgtactataatggcattggagtgcaagagattataaacaggccttgaagtatttaa 1577  
 Qy 1305 cctggcatctcagagtgggcagccctcgccatttattatctggcgaagatgtatgcaac 1364  
 Db 1578 tttagcttctcagggagggccatattctgtcttctataacctagctcagatgcagccag 1637  
 Qy 1365 aggaacagagtagtaagatcatgcagaactgctgtggagctttataaaggtgctgtga 1424  
 Db 1638 tggcagcggctgagtgcatcatgcacatgcagtgagttgtttaagaatgtatgta 1697  
 Qy 1425 actgagccactggcctgagaaattcctgcagcttactttgcttataaggtggtgatat 1484  
 Db 1698 acggggcgttggctgaaagcttatgactgctctataacagctataaagatggcgatta 1757  
 Qy 1485 agattcttcttctgttcagtgatgcactgcttgcaaaaatggggtatgaagtagctcaag 1544  
 Db 1758 caatgctgcagtgatcccagtagctcctcctggctgaacagggctatgaagtgcacaaag 1817  
 Qy 1545 caattcagcattcattttggaaattcaaaaggctaaacttcttggaaaagagaagatgta 1604  
 Db 1818 caatgcagcctttattcttgatcagagagaagcaagcattgtaggtagaagatgaaactta 1877  
 Qy 1605 tccaatggcgtctcctatggaatcgagctgccttcaaggccaatgcatttgcctagagt 1664  
 Db 1878 tcccagagcttgctacattggaaacagggcgcctctcaaggctactactgtgctagaat 1937  
 Qy 1665 aaaaattggagattaccattactatggctatgggactaagaagaagactatcaaacagcagc 1724  
 Db 1938 taagctggagactaccattctatgggttggcccgatgtagattatgaaactgcatt 1997  
 Qy 1725 cacacactacagcattgcagccaataatacaaacgcaagcgaagcatttcaactctggc 1784  
 Db 1998 tattcattacgcttggcttctgagcagcaacagagtcacaagctatgttcaactctggg 2057  
 Qy 1785 ttatatgtatgaacagcgtttagctacacaaagacatctcacttggccagagaagattga 1844  
 Db 2058 atatatgatgagaagagcagtggtgcatataaacagagattatccacttgcgaacgtttta 2117  
 Qy 1845 agacatggctgctcaaacagtcagatgccacataactgctgctcttggcgtcatgaa 1904  
 Db 2118 tgacatggcagctgagcgcagccagatgcacaagctccagctctcttagcctctgcaa 2177  
 Qy 1905 actggaactacgacttggctcgggatatctgttttataatcagtttcaaacagagatg 1964  
 Db 2178 attgggctgctctcttctgtcagtagacatcagg---gaaacaaacatctcgagatgct 2234  
 Qy 1965 gaactggctgaaactggacaacacacattggaccacactgggagcttatttggatggcct 2024  
 Db 2235 caccacacttgataggaccagcttgggacgtgagtgaggacatttaccatcatgacct 2294  
 Qy 2025 cattg 2029

```
Db 2295 cattg 2299
|||||
RESULT 9
AAX76577
ID AAX76577 standard; cDNA; 2382 BP.
XX
AC AAX76577;
XX
DT 10-AUG-1999 (first entry)
XX
DE Human pancreas-specific tumour suppressor TSA305 gene SEQ ID NO:2.
XX
KW Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;
KW therapy; tumour suppressor; ss.
XX
OS Homo sapiens.
XX
PN WO928457-A1.
XX
PD 10-JUN-1999.
XX
PF 25-NOV-1998; 98WO-JP05306.
XX
PR 20-APR-1998; 98JP-0126803.
XX
PR 28-NOV-1997; 97JP-0343789.
XX
PA (SAKA ) OTSUKA PHARM CO LTD.
XX
PI Harada Y, Ozaki K;
XX
DR WPI; 1999-358128/30.
XX
DR P-PSDB; AAX17750.
XX
PT Tumour suppressor gene TSA305 expressed specifically in pancreas,
PT useful for diagnosis and treatment of pancreatic cancer
XX
PS Claim 2; Page 42-43; 54pp; Japanese.
XX
CC The present sequence represents a gene designated TSA305, which is
CC expressed specifically in pancreas tissue. The gene, sequences
CC hybridizing with it, its expression product, and antibodies recognizing
CC the expression product, are useful in the investigation, diagnosis,
CC prevention and treatment of pancreatic cancer.
XX
SQ Sequence 2382 BP; 689 A; 501 C; 614 G; 578 T; 0 other;

Query Match 30.5%; Score 631.4; DB 20; Length 2382;
Best Local Similarity 60.5%; Pred. No. 5.3e-142;
Matches 1056; Conservative 0; Mismatches 686; Indels 3; Gaps 1;

QY 285 acaaaagcagcagagaaattttacagatgaagagagaccagctatttaagatgggaat 344
DB 513 agaaagcgtgtaagagacgcagatgcaggaagcagaatgatgtatcaaaactggaa 572
QY 345 caaggtctccagcagctcaaaagccaaacaaagaaagagagcctactcttttgc 404
DB 573 gaaatccttaatggaagcaataaagaaagccaaagaaagagagacatcgttatccca 532
QY 405 caaagcagctgacatgggaacttgaagctatggagaaatggctgacgttttgcatt 464
DB 633 aaaggcagaagcatgaaccataccaaagccctggagagtgatcgtctttatt 692
QY 465 tggaaatttggcgtgcaaaataaacagcagctatccaaattatagtccttggttaa 524
DB 693 tggatattacttgcacagaataccagcagcagagagatgtttggaagctgactga 752
QY 525 aqaagatcatgaagcccaaacgcatagatttttcttcttcttgggaataggaat 584
DB 753 ggaagcctcccaaggacagactgctcttggcttcttcttcttcttcttcttggagt 812

QY 585 ggaatatgatcaagctaaaggcactgatattacacaccttggaaagtgtgggaaacat 644
DB 813 taattcaagtcaggcgaagcctctgttatattatatactttggagctctggggcact 872
QY 645 gatgccagatgatttggggtagacagatatttgcgggaatcaatgttctacagaattg 704
DB 873 aatagccacatggtttgggttacagactgggtggtggtggtggtggtggtggtggtg 932
QY 705 tgaagtgtccctaaattattacaagaagtgccagattatattgctgacacattgaaaa 764
DB 933 tgaatctgccctgactcactatctgtgttgcacatcatgtgtgtagatatactcgt 992
QY 765 aagtgaaggtgttccagtggaagagtgagactaaacggaagacacctgaaatctgagttc 824
DB 993 aacaggaggtcagtagtacagagaatacggctgctgatgaatgggaaatcccaggaat 1052
QY 825 taacagtgagatttggattgggacatatataccaaactataaatttttggcagaagagg 884
DB 1053 gaacagtggaatgtagaagaagatttattcaattaccagttctctagctgaaaaag 1112
QY 885 agatgttcagatacaagctctcttggacaattacatctaatggcaggaaggtctaga 944
DB 1113 tgatgtacaagcacaggttggcttggacaactgcacctgcacggagggcgtagtaga 1172
QY 945 tcaggattactacaagcattacactacttcttaaggcagcaagggcgagtgcaaa 1004
DB 1173 acagaatcatcagagacatttgactacttcaatttagcagcaaatgctggcaattaca 1232
QY 1005 tgcctggcattatagaagaatgatttagaggggaatgctgcgtgcccacaaataa 1064
DB 1233 tgcctggcatttgggaagatgattcgggaagagtgacattgacctcagagtaa 1292
QY 1065 cgctactgcttcaagtagcttctccatggcagcagtaggaagggcaatcgcccttca 1124
DB 1293 tgagacagctctccactactttaagaagctgctacatggcaaccagctggacagag 1352
QY 1125 tgggttggcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1184
DB 1353 tgggttggaaaggcctacaccttcttcttcttcttcttcttcttcttcttcttctt 1412
QY 1185 taatcttctcagaagctgcggaaggggggggggggggggggggggggggggggggggg 1244
DB 1413 taagtatttccagaagctgctggaagggctgggtggtggtggtggtggtggtggtggt 1472
QY 1245 catgtactacttgcgtctggaatggaagattataaacttgccttcaaatattttta 1304
DB 1473 catgtactataatggcattggagtcagagagattataaacaggcctggaagtattta 1532
QY 1305 cctggcatctcagatggggcagccctcgcatttatttcttcttcttcttcttcttctt 1364
DB 1533 tttagcttctcagggaggccatcttcttcttcttcttcttcttcttcttcttcttct 1592
QY 1365 aggaacagagtagtaagatcatcagaaactgctgagctttataaaagggtctgtga 1424
DB 1593 tggcaccggcgtgagcagatcatcactcagtcagtgagtggtttaaagaatgtatga 1652
QY 1425 actagggcactggcgtgagaaattcctgacagcttacttgcctataagagtggtgat 1484
DB 1653 acgagccgttggctgaaaggcttactgcctataacagctataaagatggcgatta 1712
QY 1485 agatcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1544
DB 1713 caatgctgcagtgatccagtaacctcctcctggctggaacgggctgataagtggtg 1772
QY 1545 caattcagacttcttggaaatcaaaaggctaaacattcttgaagagagaatgta 1604
DB 1773 caatgcagccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1832
QY 1605 tccaatggcgtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1664
DB 1833 tcccagagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1892
QY 1665 aaaaattggagattaccattacttcttcttcttcttcttcttcttcttcttcttctt 1724
```



Db 1254 gaacagtgaatgctagagaagatttgatttcaattattaccagtttccctagctgaaaaagg 1313  
QY 885 agatgttcagatacaagttctcttggacaattacaatcattggcaggagaaggtctaga 944  
Db 1314 tgatgtacaagcacaggttggtcttggacaactgaacctgcagggcggtggagtaga 1373  
QY 945 tcaggattactacaagcattacactactctttaaaggcagcaaaagccgggagtgcaaa 1004  
Db 1374 acagaatcatcagagagcatttgactacttcaatttagcagcaaaatgctggcaattcaca 1433  
QY 1005 tgccatgcatattatagaaaagatgtatttagaggggaatgctgcogtgcgcacaaataa 1064  
Db 1434 tgccatggtccttttgggaagatgtattcggaaagatgacattgtacctcagagtaa 1493  
QY 1065 cgctactgcttcaagtacttttccatggcagccagtaaggaatgcaatcggtccttca 1124  
Db 1494 tgagacagctctccactactttaagaagctgctgacatgggcaaccacagttggacagag 1553  
QY 1125 tgggcttggtcttcttacttctaatgaaagagttccctgaattatgccraagcact 1184  
Db 1554 tgggcttggaatggcctactctctatggagaggaggtcaagttaattatgatctagccct 1613  
QY 1185 taaatcttccagaagctgcgaaaaaggtggcccgacgcacagttccagttaggctt 1244  
Db 1614 taagtattccgaaaagcgtcgtgaacaagcctgggtggatggcgacgtacagcttggttc 1673  
QY 1245 catgtactactggtcctcgtgaatatggaaggattataaacttgccttcaaatatttta 1304  
Db 1674 catgtactaatgcatggagtgcaagagagattataaaaggccttgaagtattttaa 1733  
QY 1305 cctggcatctcagtgaggcagccctcgccttcttctataacctagctcagatgcagccag 1793  
Db 1734 tttagcttctcagggagggccatctctgttcttataacctagctcagatgcagccag 1793  
QY 1365 aggaacagagtagtaagaatcatgcagaactcgtgtggagctttataaagggtgtctgtga 1424  
Db 1794 tggcacggcgtgatgcgatgtcgcacactcagtcgaggtgtgttaagaatgtatgtga 1853  
QY 1425 actaggccactgggtgagaatactctgcagcttacttgcctataaggatggtgatat 1484  
Db 1854 acgagccgttggttgaaagcttatgactgcctataaacagctataaagatggcgatta 1913  
QY 1485 agattctctctgttcagatgcaactcgttgcagaaatgggggtatgaagtgcctcaag 1544  
Db 1914 caatgctgcagtgatccagctactcctcctggtcgtgaacagggcgtatgaagtggcacaag 1973  
QY 1545 caatcagcatcatttggatctaaagaagcctaacttctgaaagagaagatgta 1604  
Db 1974 caatgcagccttattcttctgacagagaagcaagcattgtagtgagaatgaaactta 2033  
QY 1605 tccaatggcgcttctctatggaatcgagctgcccattcaagcaatgcatttgcctagagt 1664  
Db 2034 tcccagagcttggctacattggaacggcgccctcctaagctataactgtgctagaat 2093  
QY 1665 aaaaattggagattaccattactatggctatgggactaagaagaactatcaaacagcagc 1724  
Db 2094 taagctcgagactaccatttctatgggttggccaccgatgtagattatgaaactgcatt 2153  
QY 1725 cacacactacagcattgcagccaacaaataccacaacgcgaagcattgttcaatctggc 1784  
Db 2154 tattcaattaccgtctggtctctgacagcaacacagctgacaaagctatgttcaatctggg 2213  
QY 1785 ttatatgtatgaacacggcttaggcatacacaaggacattcacttggccagagaattgta 1844  
Db 2214 atatatgtatgaagagactgggtgataaacaagcagattatccacttgcgaacgttttta 2273  
QY 1845 cgacatggtcgtcacaagagtcagatgcccacataactctgtgctcttggccgtcatgaa 1904  
Db 2274 tgacatggcagctgaagccagcccagatgcacaagttccagttcttctagcctctgcgaa 2333  
QY 1905 actggaataactacgcatctggtcgggatactcctgttttttaactagttcacaacagatg 1964  
Db 2334 attggcgctcgtctattcttctgcatcacatcgg---gaaacaaacattcagatgttt 2390

QY 1965 gaactggtgaaactggacaacacacattggaccacactggaggacttattgtgtatggcct 2024  
Db 2391 caccacacttgataggaccagcttttgggaactgagtgaggacatttaccctcatgacct 2450  
QY 2025 cattg 2029  
Db 2451 cattg 2455

RESULT 11  
AAAX77013  
ID AAAX77013 standard; DNA; 3666 BP.  
XX  
AC AAAX77013;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Mouse Sel-1L splice variant coding sequence.  
XX  
KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;  
KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;  
KW vascular dementia; Parkinson's disease; coronary heart disease;  
KW fat metabolism; cholesterol metabolism; ss.  
OS Mus sp.  
PN WO9927088-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 19-NOV-1998; 98WO-CA01058.  
XX  
PR 28-JUL-1998; 98US-0123549.  
PR 19-NOV-1997; 97US-0066140.  
XX  
PA (MOUN ) MOUNT SINAI HOSPITAL.  
XX  
PI Bernstein A, Donoviel D;  
XX  
DR WPI; 1999-357833/30.  
DR P-PSDB; AAY18096.  
XX  
PT New Sel-1L nucleic acid molecule useful in the treatment of  
PT Alzheimer's disease, diabetes and cancer  
XX  
PS Claim 2; Page 71-73; 77pp; English.  
XX  
CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously  
CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and  
CC the host cell containing it can be used to prepare a Sel-1L protein.  
CC Compositions containing the Sel-1L proteins, or peptides that interfere  
CC with their binding can be used in a method for treating or preventing  
CC Alzheimer's disease, diabetes (especially insulin dependent diabetes  
CC mellitus), cancer (especially pancreatic cancer), stroke, vascular  
CC dementia, Parkinson's disease, or coronary heart disease. The  
CC compositions can also be used to treat conditions requiring modulation of  
CC fat or cholesterol metabolism.  
XX  
SQ Sequence 3666 BP; 956 A; 881 C; 953 G; 874 T; 2 other;

Query Match 28.8%; Score 596.4; DB 20; Length 3666;  
Best Local Similarity 60.0%; Pred. No. 1.6e-133;  
Matches 1031; Conservative 0; Mismatches 682; Indels 6; Gaps 2;  
QY 314 atgaaggagaccagctatttaagatgggcatacaaggttctccagcagctctaaaagccaaa 373  
Db 455 aggaagcagagatgatctatcatcaggccgggataagatactgaatgaagcaataggaaaga 514  
QY 374 aacaaaaagaagaagccttaccctactcttttgcgaagcagctgacatgggaacttgaaag 433  
Db 515 gccaaaaagagaagcagctatcgtgtaccttccgaaggcagcgagcgaatcaccaccaag 574



Claim 2; Page 66-69; 77pp; English.

PS This sequence encodes a Sel-1L (Sel-1 like) protein (also previously  
XX known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and  
CC the host cell containing it can be used to prepare a Sel-1L protein.  
CC Compositions containing the Sel-1L proteins, or peptides that interfere  
CC with their binding can be used in a method for treating or preventing  
CC Alzheimer's disease, diabetes (especially insulin dependent diabetes  
CC mellitus), cancer (especially pancreatic cancer), stroke, vascular  
CC dementia, Parkinson's disease, or coronary heart disease. The  
CC compositions can also be used to treat conditions requiring modulation of  
XX fat or cholesterol metabolism.

SQ Sequence 3667 BP; 956 A; 881 C; 954 G; 874 T; 2 other;

Query Match 28.7%; Score 594.8; DB 20; Length 3667;

Best Local Similarity 59.9%; Pred. No. 3.8e-133;

Matches 1030; Conservative 0; Mismatches 683; Indels 6; Gaps 2;

QY 314 atgaaggagaccagctatttaagatgggcatcaagggttcaccagcagctcaaaagccaaa 373

DB 455 aggaagcagagatgactatcagccgggatgaagatactgaatggaagcaatagaaaga 514

QY 374 acaaaaagaagaagcctacctactttttgccaagcagctgacatgggaagacttgaag 433

DB 515 gccaaaagagaagcatatcggtaccttcagaaggcagcagcagcgaatcacacaaag 574

QY 434 ctatggagaaatggctgacgtttgctatttggaaattttggcgtgcaaaatataaag 493

DB 575 cctggagagagtgctctatgctctcttctgttgggtattacctcacacagaataccagg 634

QY 494 cagctatccaattatagtgcttggtcaaaagagcagtcagtaaaagcccaaaagcga 553

DB 635 cagcgaagagatgtttgagaactgactgaggaaaggggtctcccaagagcagactggct 594

QY 554 taggattttgtcttcttgaataggaatggaatgatcaagctaaagcagctgatat 613

DB 695 ttggctttctacgcttctggcttgggtgttaattcaagtcaggcgaaggtcttctgtat 754

QY 614 attacaccttgaagtgctgagaaacatgatgtccagatgatttgggtacagat 673

DB 755 attactcttgagctcttgaggcaacctgatgaccatgatatttgggtaccgct 814

QY 674 atttgcgggaatcaatttctacagaattgtgaagtgcctcaagtattattacaagaaag 733

DB 815 actggctggcagtgagctcccgaggttgagtcggcactgaccatattctgtt 874

QY 734 tggcagattatattgctgacacatttgaaaaaagtgaaaggttccagtggaagaaagtg 793

DB 875 ttgcaatcatgttgctagtgatatctcctcaactggaggctctgtagtcagagaatac 934

QY 794 gactaacggaagacctgaaatctgagttctaacagtgagatttgggtggacatat 853

DB 935 ggtcgcggatgaagtggaaaaccggggatgaacagtggggtctggaagaagacctga 994

QY 854 accaattactataaatttggcagaaaagagagatgttcagatacaagtctctcttggac 913

DB 995 tticagtattaccagttctagctgagaaggtgagctccaaacacaggttggcctgggac 1054

QY 914 aattacatctaattggcaggaaggtctagatcaggattactacaagaacattacactact 973

DB 1055 agctgcatctgcatggagcgctggagtgaagaagaataccacagagcgcttggactact 1114

QY 974 tcttaaaagcagaagccggagtgcaaatgccatggcatttataggaagatgtatt 1033

DB 1115 tcaacttagcagaatactgccaattccacatgctatggcctctctctgggaagatgtatt 1174

QY 1034 tagaggggaatgctgcgtgcgcgcaaaaataacgctactctctcaagtactttccatgg 1093

DB 1175 ctgaaggaagtgcacatcgtaacctgagatgaatgagacgcgacttcaactacttaagaa 1234

QY 1094 cagccagtaagggaatgcaatcgcccttcattgggttcttcttacttacttcaatgga 1153

DB 1235 ctgctgacatgggcaaccocgtgggacagagcgggcttggaatggcctacctctacggaa 1294

QY 1154 aaggagttcccttgatattatgacraagcaacttaatactttcagaaaagctgcggaaaaag 1213

DB 1295 gagggttcaagttaattatgacctggcctcaagtatttccagaaaagctgctgagcaag 1354

QY 1214 ggtggccagcagacagcttccagtttaggttcatctactctgctctggaatagga 1273

DB 1355 gctgggtgagcggcagctgcagctcgctctatgtactacaatggcattggagtcaga 1414

QY 1274 aggattataaacttgcctcaaatatttttacctggcactcagagtgggcagccctcg 1333

DB 1415 gagattaaagcagccttggaagtattttaactgtgcttccaaggagccatattctgg 1474

QY 1334 ccatttattatctggc---caagatgtatgcaacaggaacagagtagtaagatcatgca 1390

DB 1475 cttttataacctgcacnnaagatgcacgcagcgcagcgggtgatgcgctctgctgc 1534

QY 1391 gaactgctgtggagcttttataaagggtgctgtaactaggccactgggctggaatactcc 1450

DB 1535 acactgcagtgaggtgtttaagaatgtgtgagcggaggtcgctggctcagagagactga 1594

QY 1451 tgacagcttactttgcttataaggatggtgatagattcttcttcttcttcttcagtatgac 1510

DB 1595 tgactgctacacagctataaggatgagactacaatgctgagtggtccagtaacctcc 1654

QY 1511 tgcttgcgaaaaatgggtgatgaatgactcaagcaattcagcattcatttggaaatcta 1570

DB 1655 tgcttgctgagcagcgtcacggtggcgagcaagcagccttcatcctcgcagcaga 1714

QY 1571 aaaaagcttaacattcttgaagaagagaagatgtatcccaatggcgtctccttatgaaatc 1630

DB 1715 gagaagcaacctatgagtgagaatgaaacttaccacagacttactgcatggagaaca 1774

QY 1631 gagctgccaattcaaggccaatgcatgtgtagagtaaaaaattggagattaccattactatg 1690

DB 1775 gggcgccctcccaaggttacactgtggctagaaattaagcttgagactaccactctatg 1834

QY 1691 gctatggactaagaagaactatcaaacagcagccacacacacacacacacacacacac 1750

DB 1835 gcttggcactgatgtggattatgagccgcatatttattcattaccgctggtcttgagc 1894

QY 1751 aatacccaaacgcgaagccatttcaatctggcttatatgtatgaacacgcttaggca 1810

DB 1895 agcagcagcgcgccaagctatgtttaacctgggtctacatgcacgaaaggcttaggca 1954

QY 1811 tcacaaaagacattcacttggccagaagattgtacgacatgctgctcctcaaacagctccag 1870

DB 1955 ttaaacagagacattcacttggcaaacgcgttttatgacatggcagccggaagctagccag 2014

QY 1871 atgcccacatactgtgctcttgcgttcacatgaaactggaaactacgcaatttgcctcggg 1930

DB 2015 atgcacagtaactgtgttccctgcactctgcaaatagggtgtgctctatttctacagt 2074

QY 1931 atatcctgttttttaactcagttccaaacagagatggaactggcgtgaaactggcaacacca 1990

DB 2075 acatacgggaagcaa---acattcgagatctatttcacacacatgatatggaccagcttt 2131

QY 1991 ttggaccacactgggacttatttggattggcctcattg 2029

DB 2132 tgggaccggagtgaggacctttaccctcatgaccatcattg 2170

RESULT 13

AAF93617

ID AAF93617 standard; cDNA; 557 BP.

XX AAF93617;

XX AAF93617;

DT 21-MAY-2001 (first entry)

XX cDNA encoding SRT protein isolated from testis tissue SEQ ID 438.

XX	Human; SRT; gene therapy; gene mapping; tissue typing; ss.
KW	
XX	Homo sapiens.
OS	
PN	WO200107611-A2.
XX	
PD	01-FEB-2001.
XX	
XX	21-JUL-2000; 2000WO-US200006.
PR	
PR	26-JUL-1999; 99US-0145701.
XX	
XX	(GETH ) GENENTECH INC.
PA	
XX	Baker KP, Goddard A, Wood WI;
PI	
XX	WPI; 2001-112729/12.
XX	
XX	New isolated nucleic acid molecule encoding a SRT polypeptide is useful
PT	for production of recombinant SRT polypeptides, gene mapping,
PT	diagnosing genetic disorders and for gene therapy -
PT	
XX	
PS	Claim 2; Fig 438; 663pp; English.
XX	
CC	Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC	human SRT proteins. The cDNA sequences are isolated from various
CC	different human tissue cDNA libraries. The invention relates to a method
CC	for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC	encoding SRT, a host cell transformed with the vector, an isolated SRT
CC	polypeptide, and an antibody which binds to SRT. The polynucleotide
CC	sequence can be used in gene therapy and is useful in the recombinant
CC	production of SRT polypeptides, as a hybridisation probe to screen
CC	libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC	map the gene encoding the SRT polypeptides and analysing genetic
CC	disorders, tissue typing and disease tissue detection. The SRT
CC	polynucleotide sequences can be used in polymerase chain reaction,
CC	screening for new therapeutic molecules and generation of antisense RNA
CC	and DNA.
XX	
SQ	Sequence 557 BP; 212 A; 81 C; 120 G; 142 T; 2 other;

Query Match	24.7%	Score	510.4	DB	22	Length	557
Best Local Similarity	97.7%	Pred. No.	3.6e-113				
Matches	549	Conservative	0	Mismatches	8	Indels	5
Gaps							

  

Qy	206	agaaaaagaagaatcaacgtaaaataagaataaaaaagaattcaaaaaataaagatatcttga	265
Db	1	agaaaaagaagaatcaacgtaa---ataagataaaggattcaaaaaataagatntctga	57
Qy	266	agaaaaataagaatcatttacaagaacaagcagagaaaaattttcacagatgaaggagacc	325
Db	58	agaaaaataagaatcatttacaagaac--agcagagaaaaattttcacagatgaaggag--cc	115
Qy	326	agctatttaagatgggcatcaaggtttctccagcagctctaaagccaaacaaacaaagaag	385
Db	116	agctatttaagatgggcatcaaggtttctccagcagctctaaagccaaacaaacaaagaag	175
Qy	386	aagcctacctacttttgcacaaagcagctgcacatgggaaacttgaagcctatgagaaaaa	445
Db	176	aagcctacctacttttgcacaaagcagctgcacatgggaaacttgaagcctatgagaaaaa	235
Qy	446	tggctgacgcttgcctattttggaattttggctgcacaaatataacacgacgtatccaat	505
Db	236	tggctgacgcttgcctattttggaattttggctgcacaaatataacacgacgtatccaat	295
Qy	506	tatatgactcttggctaaagaagatcatgtaaagcccaaacgcataggatttttgt	565
Db	296	tatatgactcttggctaaagaagatcatgtaaagcccaaacgcataggatttttgt	355
Qy	566	cttcttatggaataggaaatggaatatgatcaagctaaagcgaactgatatattacaccttg	625

Db	356	cttcttatggaataggaaatgatgaactgaagcactgatatattacacctttg	415
Qy	626	gaagtgcgtggaggaaacatgatgtccccagatgattttgggggtacagatatttgcgggaa	685
Db	416	gaagtgcgtggaggaaacatgatgtccccagatgattttgggggtacagatatttgcgggaa	475
Qy	686	tcaatgttctcacagaattgtgaagtgtccctaaagtatttacagaagaagtcgagattata	745
Db	476	tcaatgttctcacagaattgtgaagtgtccctaaagtatttacagaagaagtcgagattata	535
Qy	746	ttgtgcacacatttgaaaaag	767
Db	536	ttgtgcacacatttgaaaaag	557
RESULT 14			
AAF93859			
ID	AAF93859	standard; cDNA; 2144 BP.	
XX	AC	AAF93859;	
XX	DT	23-MAY-2001 (first entry)	
XX	XX	Human cDNA encoding a membrane or secretory protein clone PSEC0209.	
XX	DE	Human; secretory protein; membrane protein; vaccine; gene therapy;	
KW	KW	rheumatoid arthritis; diabetes; ss.	
XX	OS	Homo sapiens.	
XX	PN	EP1067182-A2.	
XX	PD	10-JAN-2001.	
XX	XX	07-JUL-2000; 2000EP-0114090.	
PF	PF	08-JUL-1999; 99JP-0194179.	
XX	PR	11-JAN-2000; 2000JP-0118775.	
PR	PR	02-MAY-2000; 2000JP-0183766.	
XX	XX	(HELI-) HELIX RES INST.	
PA	PA	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;	
PI	PI	WPI; 2001-093989/11.	
DR	DR	P-PSDB; AAB88432.	
XX	XX	Nucleic acids encoding secretory proteins/membrane proteins, useful in	
PT	PT	gene therapy or as candidate target molecules in drug development -	
XX	XX	Claim 1; SEQ ID 231; 609pp + CD ROM; English.	
PS	PS	This invention relates to nucleic acid sequences AAF93744 - AAF93916	
XX	XX	which encode human secretory or membrane proteins represented by	
CC	CC	AAB88817 - AAB88419. Included in the invention are primers	
CC	CC	AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the	
CC	CC	cDNA sequences of the invention. The invention also includes methods for	
CC	CC	the production of antibodies directed against the proteins, and cDNA	
CC	CC	sequences, which can be used in vaccines. The polynucleotide sequences	
CC	CC	can be used in gene therapy. The polynucleotide sequences and the	
CC	CC	proteins they encode may be used in the prevention, treatment and	
CC	CC	diagnosis of diseases associated with inappropriate secretory	
CC	CC	protein/membrane protein expression. The nucleic acids and complementary	
CC	CC	sequences may also be used as DNA probes in diagnostic assays	
CC	CC	(e.g. polymerase chain reactions (PCR) to detect and quantitate the	
CC	CC	presence of similar nucleic acid sequences in samples. They may also be	
CC	CC	used to study the expression and function of secretory proteins/membrane	
CC	CC	polypeptides and their role in metabolism. The polypeptides may be used	
CC	CC	as antigens in the production of antibodies against them and in assays to	
CC	CC	identify modulators (agonists and antagonists) of expression and	
CC	CC	activity. The antibodies and antagonists may also be used as therapeutic	
CC	CC	agents to down regulate expression and activity. The antibodies may also	
CC	CC	be used as diagnostic agents for detecting the presence of the	

CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
CC (ELISA). Examples of diseases which may be treated include rheumatoid  
CC arthritis and diabetes.  
XX  
SQ Sequence 2144 BP; 623 A; 407 C; 552 G; 562 T; 0 other;

Query Match 21.4%; Score 444; DB 22; Length 2144;  
Best Local Similarity 60.4%; Pred. No. 5,1e-97;  
Matches 732; Conservative 0; Mismatches 480; Indels 0; Gaps 0;  
QY 285 acaaaagcagcagagaaattttacagatgaagagaccagctattttaagatggcat 344  
DB 609 agaagaggtgctaaagacgagcagatgaggaagcagaaatgatgaacaaactggaat 668  
QY 345 caaggtctccagcagctaaagccaaacaaacaaagaaagaaagcctactacttttgc 404  
DB 669 gaaatccttaagtgaagcaataagaaagccaaagaaagaaagcagatcgttatccca 728  
QY 405 caagcagctgacatgggaacttgaaagctatggagaaatggctgacgtttgctatt 464  
DB 729 aaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 788  
QY 465 tggaaattttggcgtgcaaaataataacacagcagcagcagcagcagcagcagcagc 524  
DB 789 tggatgattcttgccacagaaatccagcagcagcagcagcagcagcagcagcagcag 848  
QY 525 agaagatcatgcaaaagcccaaaagcagcagcagcagcagcagcagcagcagcagcag 584  
DB 849 gaaagcgtcccaagggcagcagcagcagcagcagcagcagcagcagcagcagcagc 908  
QY 585 gaaatgatcaagctaaagcagcagcagcagcagcagcagcagcagcagcagcagcag 644  
DB 909 taattcaagtcaggcaaaagcgtctgtatattacatttggagcgtcttggggcaatct 968  
QY 645 gatgtccagatgatttgggtacagatatttggggaacaaatgtctctacagaattg 704  
DB 969 aatagccacatgggtttgggttacagatcactgggtggcagcagcagcagcagcag 1028  
QY 705 tgaagtgccctaaagtattacaagaaagtgccagcagcagcagcagcagcagcagcag 764  
DB 1029 tgaatcgtccctgactcactatcgtctgttgccaatcagttgctagtgatctcgt 1088  
QY 765 aagtgaaagtggttccagtggaagtgagcagcagcagcagcagcagcagcagcagcag 824  
DB 1089 aacaggaggtcagtagtacagagaatacgcgtgctgatgaagtggaaatccaggaat 1148  
QY 825 taacagtggagatttggattggagacataaccataactataaatttttggcagaagagg 884  
DB 1149 gacagtggaatgctagaagagatttgattcaattaccaggttccctagctgaaaaagg 1208  
QY 885 agatgttcagatcacagctctctctggacaattacatctaaatggcaggaagagcttaga 944  
DB 1209 tgatgtacaagcacaggttggctctggacaactgcactgcagcagggcgtggagtaga 1268  
QY 945 tcaggattactcaaaagcattacactactctctaaaggcagcaagggcggagtgcaaa 1004  
DB 1269 acagaatcatcagagagcatttgactacttcaatttagcagcaaatgctggcaattcca 1328  
QY 1005 tgccatgcatattataggaagatatttagagggaatgctgcgtgcgcgcaaaataa 1064  
DB 1329 tgccatggccttttgggaagatgattctggaaagagtgacattgacctcagagtaa 1388  
QY 1065 cgctactgcttcaagcattttccatggcagccagcagcagcagcagcagcagcagcagc 1124  
DB 1389 tgagacagctctccactactttaagaagctgctgacatgggcaacccagttggcagag 1448  
QY 1125 tgggtgtgctcttcttacttcaatggaaagaggttcccttgaattatgcraagcact 1184  
DB 1449 tgggtgtggaatggcctactctatgggagagaggttcaagttaattatgatctagcct 1508  
QY 1185 taaacttctcagaagctcggaagaaagggtggccgacgacagtcacagttccagttagg 1244

DB 1509 taagtattccagaaagctgctgaacaaagctgggtggatgggcagctacagcttggttc 1568  
QY 1245 catgtactactctgctctggaatggaatggaagattataaactgcttcaaatatttta 1304  
DB 1569 catgtactataatggcattggagtcgaagagagattataaacacagggccttgaattttta 1628  
QY 1305 cctggcatctcagatgggcagccccctccttatttattctctggccaagatgtatgcaac 1364  
DB 1629 tttagcttctcagggagggccatatcttggctttctataaccatgctcagatgcagccag 1688  
QY 1365 aggaacagagagtagtaagatcatgcagaactgctgagagctttataaaaggtgtctgtga 1424  
DB 1689 tggcaccggcgtgatgcgatcatgtccactgcagtgagggttttaagaatgtatgtga 1748  
QY 1425 actagccactggcgtgagaaattctcagacttacttgcctataagagatggtgat 1484  
DB 1749 acgagggcgttggctgaaagcctttagctccataacacagctataaagatggcgatta 1808  
QY 1485 agattctctct 1496  
DB 1809 caactactgctg 1820  
RESULT 15  
AAAX77014  
ID AAAX77014 standard; DNA; 2109 BP.  
XX  
AC AAAX77014;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human Sel-1L coding sequence.  
XX  
KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;  
KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;  
KW vascular dementia; Parkinson's disease; coronary heart disease;  
KW fat metabolism; cholesterol metabolism; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09927088-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 19-NOV-1998; 98WO-CA01058.  
XX  
PR 28-JUL-1998; 98US-0123549.  
PR 19-NOV-1997; 97US-0066140.  
XX  
PA (MOUN ) MOUNT SINAI HOSPITAL.  
XX  
PI Bernstein A, Donoviel D;  
XX  
DR WPI; 1999-357833/30.  
DR P-PSDB; AAY18097.  
XX  
PT New Sel-1L nucleic acid molecule useful in the treatment of  
XX Alzheimer's disease, diabetes and cancer  
PS Claim 2; Page 75-76; 77pp; English.  
XX  
CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously  
CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and  
CC the host cell containing it can be used to prepare a Sel-1L protein.  
CC Compositions containing the Sel-1L proteins, or peptides that interfere  
CC with their binding can be used in a method for treating or preventing  
CC Alzheimer's disease, diabetes (especially insulin dependent diabetes  
CC mellitus), cancer (especially pancreatic cancer), stroke, vascular  
CC dementia, Parkinson's disease, or coronary heart disease. The  
CC compositions can also be used to treat conditions requiring modulation of  
CC fat or cholesterol metabolism.  
XX  
SQ Sequence 2109 BP; 591 A; 442 C; 492 G; 584 T; 0 other;



Query Match					
Best Local Similarity		21.1%	Score 437.6;	DB 20;	Length 2109;
Matches 690; Conservative		0;	Mismatches 399;	Indels	3; Gaps
Qy	938	gtctagatcaggattactacaagaacattacactacttctttaaaagcgacgaacgcccggga	997		
Dy	67	gtcgcaaacagaatcatcagagacgatttgactactccaatttagcagcaaatgtcgga	126		
Qy	998	gtgcacaaatgccatggcatttatagaaaagtattttagagggaatgtccgtgccgc	1057		
Dy	127	attcacatgccatggccttttgggaaagatgtattcggaaagggaatgtacattgtacct	186		
Qy	1058	aaaaacgcctactgccttcagactcttccatggcagcgtaagggcaatgcaatcg	1117		
Dy	187	agagtaatgagacagctctccactacttaagaaagctgtgacatgggcaaccagttg	246		
Qy	1118	gccttcattgggttgtttcttactttcatcgaaaaaggaggttccccctgaattatgcr	1177		
Dy	247	gacagagtggcttggaaatggcctacctctatgggagaggagttccaagttaattatgatc	306		
Qy	1178	aagcacttaatacttctcagaagctgcgaaaaagggtgccccagcgcacattccagt	1237		
Dy	307	tagcccttaagtattccgaaagctgtcgaaacaggctgggtggatgggacgtacagc	366		
Qy	1238	tagcttcatgtactactctggctctggaaatgaaggattataaaccttgccttcaa	1297		
Dy	367	tgggttccatgtactaataatggcattggagtcgaagagattataaacaggccttgaagt	426		
Qy	1298	atttttacctggcatctcagatggggcagccctcgccatttatctgtggccaagtgt	1357		
Dy	427	attttaatttagcttctcagggaggccatatcttggcttctctaaccctagctcagatgc	486		
Qy	1358	atgcaacaggaaacagagtagtaagaatcatcgaaacictgtggagctttataaagggt	1417		
Dy	487	atgcacgtggcacccggcgtgatgcgatcatgtccacacgtggagttgtttaagaagt	546		
Qy	1418	tcctgaactlaggccactgggtcgaaatctctgacagcttacttgcctataaggatg	1477		
Dy	547	tatgtgaacgagggcgttggtctgaaggcttatgactgcctataacagctataaagatg	606		
Qy	1478	gtgatatagattctcttcttttcagatgcactgcttgcagaaatggggtatgaagttag	1537		
Dy	607	gcgattacaatgctgcagtgatccagtagctctctctgctggcgaacgggctatgaatgg	666		
Qy	1538	ctcaagcaattcagatcatttttggaaatctaaaaaggctaaacattcttgaaaaagaga	1597		
Dy	667	cacaaagcaatgcagccttatcttgcagagagaagcaagcatgtaggtgagaatg	726		
Qy	1598	agatgatccaatggcgctctctctatgaaatcgagctggccattcaaggcaatgattg	1657		
Dy	727	aaacttatcccaagagtttgctacattggaaacaggcgccgctctcaaggctatactgtgg	786		
Qy	1658	ctagagtataaatttggagattaccattactatggcctatgggactaagaagactatcaaa	1717		
Dy	787	ctagaattaagctcggagactaccatttctatggggtttggccacgattagattatgaaa	846		
Qy	1718	cagcagccacaactacagcattgcagccaaacataccaaacgcgcgaagccaatgttca	1777		
Dy	847	ctgcatttatcataccgctcggctcttgagcagcaacacagtcgcaagctatgttta	906		
Qy	1778	atctggcttatatgatgaacggcttaggcatacaaaaggacattcaacttggccagaa	1837		
Dy	907	atctgggataatgcatagaaaggactgggcatttaaacaggattatccaccttgcgaac	966		
Qy	1838	gattgtacgacatggctgtctcaaacagttccagatgcaccaataacctgtgctctttccg	1897		
Dy	967	gtttttatgacatggcagctgaagccagccagatgcacaagttccagttctctagccc	1026		
Qy	1998	tcatgaaactggaactacgcatttgcctcgggatatctctgttttttaatcagttcacia	1957		
Dy	1027	tcctgcaaatggcgctgcgtctattcttgcagtacatacagg---gaaacaaacttcqag	1083		

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:24:18 ; Search time 1193 Seconds  
(without alignments)  
18645.232 Million cell updates/sec

Title: US-09-714-882-1

Perfect score: 2070

Sequence: 1 atgaagccctgtctctgtt.....ttagaatcaccatgggtag 2070

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	608.8	29.4	3055	12	AK005023	AK005023 Mus muscu
2	304.8	14.7	421	11	BF153071	BF153071 uz99g07.y
3	252.8	12.2	630	11	BI067239	BI067239 pgfln.pk0
4	243.8	11.8	592	10	BE680423	BE680423 df79c09.y
5	220.2	10.6	343	11	BF509718	BF509718 UI-H-B14-
6	215.2	10.4	549	11	BG577404	BG577404 N115 SSH-
7	198.2	9.6	459	10	AA310028	AA310028 EST180874
8	196.6	9.5	694	10	AL596660	AL596660 DKFZp451L
9	195	9.4	632	11	BE912256	BE912256 601665066
10	170.8	8.3	932	10	BE570424	BE570424 601333322
11	168.6	8.1	492	11	BE995163	BE995163 UI-M-C60P
12	166.6	8.0	659	11	BF727590	BF727590 SWOV3MCAM

13	165.8	8.0	404	11	BF359310	BF359310 RCS-ET008
14	165.6	8.0	721	11	BG197665	BG197665 RST16894
15	165.6	8.0	763	11	BG184301	BG184301 RST3225 A
16	163	7.9	553	10	AL596788	AL596788 DKFZp451J
17	157.6	7.6	776	11	BG185764	BG185764 RST4716 A
18	154.6	7.5	366	11	BG188306	BG188306 RST7323 A
19	153.8	7.4	179	11	BG182541	BG182541 RST1417 A
20	151.6	7.3	716	13	AQ485268	AQ485268 RPCI-11-2
21	150.6	7.3	494	10	AL589253	AL589253 DKFZp451C
22	150.2	7.3	428	10	AL751663	AL751663 cn11c01.x
23	147.8	7.1	424	11	BG191061	BG191061 RST10031
24	147.2	7.1	464	11	BG201173	BG201173 RST20501
25	143.8	6.9	542	10	AL589397	AL589397 DKFZp451N
26	141.8	6.9	178	11	BG216522	BG216522 RST36102
27	141.6	6.8	174	11	BG195560	BG195560 RST14753
28	141	6.8	564	13	AQ984054	AQ984054 RPCI-23-3
29	140.8	6.8	353	11	BG184754	BG184754 RST3576 A
30	140.4	6.8	490	10	AI425637	AI425637 md58h04.Y
31	140.2	6.8	177	11	BG204271	BG204271 RST23672
32	138.8	6.7	534	11	BG880030	BG880030 1b73h09.Y
33	137	6.6	174	11	BG203754	BG203754 RST23129
34	137	6.6	344	11	W62650	W62650 md58h04.r1
35	135	6.5	622	11	BF501529	BF501529 AT16891.5
36	134.8	6.5	178	11	BG197561	BG197561 RST16806
37	134.4	6.5	595	10	AW281997	AW281997 fj60e07.x
38	133.8	6.5	362	11	BG218395	BG218395 RST38263
39	132.2	6.4	711	10	AV712962	AV712962 AV712962
40	131.8	6.4	573	10	AI552501	AI552501 mj94h06.Y
41	130.6	6.3	176	11	BG219992	BG219992 RST39784
42	130.6	6.3	591	10	AA288149	AA288149 vbl5e09.f
43	128	6.2	212	11	BG184106	BG184106 RST3024 A
44	124.4	6.0	943	11	BE912650	BE912650 601663693
45	119.8	5.8	270	10	AV265793	AV265793 AV265793

## ALIGNMENTS

RESULT	1
AK005023	
LOCUS	AK005023 3055 bp mRNA HTC 05-JUL-2001
DEFINITION	Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300016D21, full insert sequence.
ACCESSION	AK005023
VERSION	AK005023.1 GI:12836652
KEYWORDS	CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 3055)
TITLE	Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Methods in enzymology. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	2 (bases 1 to 3055)
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (bases 1 to 3055)
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome research. 10 (11), 1757-1771 (2000)  
 20530913  
 MEDLINE  
 PUBMED  
 REFERENCE

4 (bases 1 to 3055)  
 The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 5 (bases 1 to 3055)  
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,  
 Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,  
 Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,  
 Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,  
 Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,  
 Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,  
 Okido,T., Owa,C., Saito,H., Saito,R., Sakai,K., Sano,H.,  
 Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
 Tanaka,T., Teijima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
 Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 Direct Submission  
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues. First strand cDNA was primed with a primer  
 [5' GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse transcriptase  
 and subsequently enriched for full-length by cap-trapper. Second  
 strand cDNA was prepared with the primer adapter of sequence[5'  
 GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCGCCCCC 3']. cDNA was  
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:  
 XhoI. Host: SOLR.

FEATURES  
 Source  
 Location/Qualifiers  
 1..3055  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="MGI:1329016"  
 /db\_xref="MGI:1907284"  
 /clone="1300016D21"  
 /sex="male"  
 /tissue\_type="liver"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 99..2471  
 /note="putative"  
 /codon\_start=1  
 /protein\_id="BAB23750.1"  
 /db\_xref="GI:12836653"  
 /translation="MOVVRRLSLLLCAVLGSAATSDDKTNQDLSQSKSLPTDDE  
 SVKDHHTTKGVAGQIFVDEEAEEVSLQDEDSKTOQEEISFLESPNPSKTYEE  
 LKVRPVLVTEAGHGEPCHFPFLDKEDGTSDEGRGLMCATTYDKYKDEK  
 WGFETDEDAARQMOEAEMIYQAGMKILNGSRKSKREARYLQKAGMNHKALV  
 ERYTSVALLGCDYLTIQAKNEFKELTEEGSPKGTGLFVAGLGVNSQAKALV  
 YTFGALCNLTHMLTYRYKVAIGVIGVLCESALTHYRLVANHVASDILSTGCVVQ  
 RLRPDEVENPGMNSMLEDLLQYQYFLAEKGDVQAOVGLQLHLHGGRGVQNHQR  
 ARDYFNLANAGNSHMAFLKGMVSEGSDIVPOSNETALHYFKRADMPNPVQSGVLG  
 MAYLYRGQVNVNIDALKYFQKAAEQGVQDGLQSGMYNGIGVXRKYQKALYFNL  
 ASQGVQVYFLNLAQMHASGTGYMRSCTDRAVELFKNVCGRGRSRLMTAYNKDSD  
 YNOAGHYVLLLEQQEYVAGSAQAFTLDQREATIVGNETIYPRALLHWNRAASQGVY  
 YNAAQVYLLLEQQEYVAGSAQAFTLDQREATIVGNETIYPRALLHWNRAASQGVY

Db 1528 GCTGGGTGGACGGCAGCTGCTGCTGCTCTATGTACTACAAATGGCATTGGAGTCAAGA 1587  
Qy 1274 aggtattataacttgccttcaaatatttttaccctggcatctcagagtggcagccctcg 1333  
Db 1588 GAGATTATAGCAGGCCCTTGAAGTATTTATCTTGGCTTCTCAAGGAGGCCATATCTTGG 1647  
Qy 1334 ccattattatcttgcccaagatgtatgcaacagagaagaggtagtagtaagatcagagaa 1393  
Db 1648 CTTCTATAAAGCTGACAGATGACGCGCAGCGGACAGGGGTGATGGCTCTGTGCACA 1707  
Qy 1394 ctgctgtgagctttataaagtgctgtgaactaggccactggctgagaaattcctga 1453  
Db 1708 CTGAGTGGAGTGTGTTAAAGATGTGTGAGCGAGTCTGCTGTCAGAGAGACTGATGA 1767  
Qy 1454 cagcttacttgcctataagtgatgatatagattctctctgttcagtagcaactgc 1513  
Db 1768 CTGCTTACAACAGCTATGAAGATGAGGACTACATGCTTGCAGTGGTCCAGTACCTCTGC 1827  
Qy 1514 ttgcagaaatggggtatgaagtagctcaaaatcaattcagcattcatttttggaattcaaaa 1573  
Db 1828 TGGCTGAGCAGGGGTACGAGGTGGCGCAGACGACGACGCTTCTCCTCGACACAGAGAG 1887  
Qy 1574 aggtcaaatcttgcaaaagagaagatgtatccaatggcgcttctctatggaatcagag 1633  
Db 1888 AAGCAACCATTTGAGTGAAGATGAACCTTACCCAGAGCTTTACTGCTGATTTGGAACGGG 1947  
Qy 1634 ctgcccattcaaggcaatgcatttgcctagtagtaaaattggagattaccattactatggct 1693  
Db 1948 CCGCTCCCAAGGTTACACTGTGGCTAGATTAAGCTTGGAGACTTACCATTCTATGGCT 2007  
Qy 1694 atgggactaagaagactatcaaacagcagcagccacacactacagcatttcagcccaacaat 1753  
Db 2008 TTGGCACTGATGTGATGATGATGAGACCGCATTTATTCATTACCGCTGGCTCTGTGAGCAGC 2067  
Qy 1754 accaacaacgcaacccatgttcaatctgctagtagtaaaattggagattaccattactatggct 1813  
Db 2068 AGCAGCGGCCCAAGCTATGTTTAAACCTGGGCTACATGACGAGAGAGGCGCTTAGGCATTA 2127  
Qy 1814 caaagacattcacttggccagaagatgtacagacatggctgctcaaacgagtcagatg 1873  
Db 2128 AACAGGACATCACCTTGCACAAACGCTTTTATGACATGCGACGCGCAAGCTAGCCCCAGATG 2187  
Qy 1874 cccacataactgtgctcttggccgtcatgaaactggaactacgcatgttgcctcgggata 1933  
Db 2188 CACAAGTACCTGTGTCTCGCACCTCTGCACAAATTAGGTAGTGTCTCTTATTCTTACAGTACA 2247  
Qy 1934 tctgttttttaaatcagttcaacagagatggaactgctgaaactggaactggaacacaccattg 1993  
Db 2248 TACGGGAAGCAA---ACATTCGAGATCTATTTCACAACTGGATATGGACCAAGCTTTTGG 2304  
Qy 1994 gaccacactgggacttattgtgattggcctcattg 2029  
Db 2305 GACCCGAGTGGGACCTTTACCTCATGACCATCATTTG 2340

RESULT 2  
LOCUS BF153071 421 bp mRNA EST 26-OCT-2000  
DEFINITION uz99g07.y1 NCI\_CGAP\_Tel Mus musculus cDNA clone IMAGE:3809845 5' similar to TR:Q9UGD3 Q9UGD3 D384266.2 ; mRNA sequence.  
ACCESSION BF153071  
VERSION BF153071.1 GI:11034466  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 421)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

MGI:1451957  
Trace considered overall poor quality  
Seq primer: -40RP from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
source 1. 421  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3809845"  
/sex="male"  
/dev\_stage="5 months"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: testis; Vector: PCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Library constructed by Life Technologies."  
BASE COUNT 107 a 93 c 100 g 121 t  
ORIGIN

Query Match 14.7%; Score 304.8; DB 11; Length 421;  
Best Local Similarity 83.7%; Pred. NO. 2.8e-46;  
Matches 345; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 1223 acgcaactccagtagcttcatctactctgctctgctgaatggaaggattata 1282  
Db 3 ATGCACAGTTCAGCTGGGCTTCATGTACTCTCTGGCTCTGGAGTATGGAAGGATTATA 62  
Qy 1283 aacttgccttcaaatattttacctggcatctcagagtgggcagccctcgccattatt 1342  
Db 63 AGCTTGCCTTCAAAATACITTTTACTTGGCATCTCAGAGTGGGCGAGCTCTTGCATTATT 122  
Qy 1343 atctggccaaagatgtatgcaacaggaacagagtagtagtaagatcagcagaactgctgtgg 1402  
Db 123 ATCTAGCGGATGTACGCAACTGGCACTGGAGTGTGAGATCATGCATACGCGCGTGG 182  
Qy 1403 acgtttataaagtgctgtgaactagccactggcgtgagaaattcctgacagcttact 1462  
Db 183 AGCTTTCAAAAGGTGTCTGTGAACCTAGGCCACTGGGCTGAGAAATTCCTTAACGGGTCTATT 242  
Qy 1463 ttgcctataaagtgatgatagattcttctgttcagtagtgcactgcttgcagaaa 1522  
Db 243 TCGCCTATAGGATGGGACATAGACTCTTCTTATTTCAGTATGCATTGCTTGGCGAAA 302  
Qy 1523 tggggatgaagtagctcaaaacattcagcattcatttttggaattcaaaagggttaaca 1582  
Db 303 TGGGATTTGAAGCAGCTCAAAACCCATTTCGCCCTTCTCTTTGGGATTCCTCAATAGCTGAAA 362  
Qy 1583 ttcttgaagaagagagatgtatccaatggcgcttctcctatggaatgagc 1634  
Db 363 TACTGGGAAAAGAGAGATGTATACCTAGGCACATTTGCTATTGAACCGAGC 414

RESULT 3  
LOCUS BI067239 630 bp mRNA EST 15-JUN-2001  
DEFINITION pgfin.pk011.e10 normalized chicken fat cDNA library Gallus gallus cDNA clone pgfin.pk011.e10 5' similar to gb|AAD05210.1 (AF063095) SEU.LL [Mus musculus]G, mRNA sequence.  
ACCESSION BI067239  
VERSION BI067239.1 GI:14474761  
KEYWORDS EST.  
SOURCE chicken.





Email: zhonghua.zhang@mail.vanderbilt.edu  
 DNA sequencing by: Vanderbilt Medical Center DNA sequencing core laboratory. This is a fragment of gene that up-regulated by NS-398 treatment.

## PCR PRIMERS

FORWARD: 5'-CTAATACGACTCACTATAGGGC-3'

BACKWARD: 5'-TCGAGCGCGCCGCGGCAGGT-3'

Seq primer: M13 forward primer

High quality sequence stop: 549.

## FEATURES

## Source

1..549

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="SSH-HCA-U library"

/cell\_type="Colon cancer cell"

/cell\_line="HCA-7"

/note="Vector: pCR2.1-TOPO; Poly A RNA was isolated from NS-398 treated and control cells, and suppression subtractive hybridization was performed in forward and reverse directions. The adapter sequences used in the hybridization were as follows:

5'-CTAATACGACTCACTATAGGGC-3'

(adapter 1) and

5'-CTAATACGACTCACTATAGGGCAGGTGGTGGCGCCGAGGT-3' (adapter 2R).

BASE COUNT 153 a 112 c 133 g 150 t 1 others

Query Match 10.4%; Score 215.2; DB 11; Length 549;

Best Local Similarity 64.3%; Pred. NO. 7e-30;

Matches 353; Conservative 0; Mismatches 194; Indels 2; Gaps 2;

QY 1250 actactctggctcgtgaatggaaggattataaactgcctcaaatattttacctgg 1309

DB 1 ACTATATGGCATTGGATCAGAGAGATTATAACAGCCCTTGAGTATTTTATTTAG 60

QY 1310 catctcagatgggagccctcgccattattctgcccgaagtgtatgcaacaggaa 1369

DB 61 CTTCTCAGGAGGCCATATCTTGGCTTTCTATAACCTAGTCAGATGTCAGTGGCA 120

QY 1370 caggagtagtaagatcagaaactgctgtggaactttataaaagtctctgtaactag 1429

DB 121 CCGCGGTGATGCGATCATGTCACACTGCGTGGAGTGTGTTAAGAAATGTATGTGAACGAG 180

QY 1430 gccactgggctgagaaattcctgcagacttacttggcctataaaggatggtgatagatt 1489

DB 181 GCGGTTGGTCTGAAGGCTTATGACTGGCTATACAGCTATATAAGATGGCGATTACAATG 240

QY 1490 ctctctgttcaagtatgcactgcttgcagaaatggggtatgaagtctcaaaagcaatt 1549

DB 241 CTGCAGTATCCAGTACCTCTCTGCTGCTGAACAGGGCTATGAAGTGCCACAAAGCAATG 300

QY 1550 cagcattcatttggaaacttaaaaggctaaccattcttgaaaagagagatgtatccaa 1609

DB 301 CAGCCCTTATTCTTCATCAGAGAAGCAAGCATTTGATGGTGAAGTGAATCTTATCCCA 360

QY 1610 ttggccttctctatggatcgagctgccattcaaggcaatgcatctgctagagtaaaaa 1669

DB 361 GAGCTTTGCTACATTGGACAGGGCGGCTCTCAGGGTATACTGTGCTAGAAATTAAGC 420

QY 1670 ttggagattaccattactatggctatggga-ctaagaaagactatcaaacagcagccaca 1728

DB 421 TCGGAGACTACCATTTCTATGGTTTCNGCACCAGATGTAGATTATGAACATGCTATTATT 480

QY 1729 cacta-cagcattgcagcaacaataaccacaacgcgcgaacattgttcaatctggctta 1787

DB 481 CATTACCGCTGGCTTCTGAGCAGCAACACAGTGCACAAAGCTATGTTTAAATCTGGGATA 540

QY 1788 tatgtatga 1796

DB 541 TATGATCA 549

## RESULT 7

AA310028

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AA310028 459 bp mRNA EST 19-APR-1997  
 EST180874 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.  
 AA310028  
 AA310028.1 GI:1962357  
 EST.  
 human.  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 459)

Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult  
 C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White  
 O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,  
 Cline,R.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgald  
 L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A.,  
 Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,  
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
 Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,X.,  
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
 Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,  
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,  
 Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,  
 Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon  
 M.R., Rosen,C.A., Haseitine,W.A., Fields,C., Fraser,C.M. and  
 Venter,J.C.

Initial assessment of human gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)

96026280

COMMENT

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse.

## FEATURES

## Source

1..459  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):156098"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Jurkat T-cells V"  
 /cell\_type="T-lymphocyte"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 125 a 90 c 114 g 125 t 5 others

## Query Match

Best Local Similarity 9.6%; Score 198.2; DB 10; Length 459;

Matches 293; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 1166 tgaattatgccraagcacttaatacttcagaaagctcggaagagggggccgcagc 1225

DB 2 TTAATTATGATCTAGCCCTTAAAGTATTTCCAGAAAGCTGCTGAACAGGCTGGGTGGATG 61

QY 1226 cacagttccaggtcttactgtactactctggctctggaataggagattataaac 1285

DB 62 GGCAGCTACAGCTGGTTCCATGTACTATATGGCATTTGGAGTCAAGAGATTATAAC 121

QY 1286 ttgcttcaaatatttttaccctggcatctcagagtgaggagccctcgcattttatc 1345

DB 122 AGGCCTTGAGTATTTTAAATTAGTCTCTCAGGGAGGCCATATCTTGGCTTCTATAACC 181







Db 358 CTAGGCAATTAACAGGACATTCACCTTGCCAAACGCTTTTATGACATGCGACGCAAGC 417  
 QY 1863 gagtcagatgccacatactctgtctcttgcgcgtcatgaactggaactacacattt 1922  
 Db 418 TAGCCAGATGACAGACAGTACCTGTCTCCACACTCTGCAAAATAGGTGTGCTATT 477  
 QY 1923 gctccgggatccctgttttttaacagttcacagatggaactggctgaaactgga 1982  
 Db 478 CTTACAGTACATACGGGAAGCAA---ACATTCGAGATCTATTACACAACTGGATATGGA 534  
 QY 1983 caacaccattggacacacactgggaacttattgtgattggcctcattgttct 2034  
 Db 535 CCAGCTTTTGGGACCGAGTGGGACCTTTTACCTCATGACCATCATTTGACCGT 586

## RESULT 11

BE995163/c 492 bp mRNA EST 05-OCT-2000  
 LOCUS UI-M-CG0p-b11-h-01-0-UI.s1 NIH\_BMAP\_Ret4\_S2 Mus musculus cDNA clone  
 DEFINITION UI-M-CG0p-b11-h-01-0-UI 3', mRNA sequence.

ACCESSION BE995163  
 VERSION BE995163.1 GI:10679134

KEYWORDS EST.  
 SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 492)  
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

## MEDLINE

COMMENT 9704477

Contact: Chin, H  
 National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENERICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA=No.

## FEATURES

source

Location/Qualifiers

1..492

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-CG0p-b11-h-01-0-UI"

/clone\_lib="NIH\_BMAP\_Ret4\_S2"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT7T3P-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The

NIH\_BMAP\_Ret4\_S2 library is a subtracted library,

ultimately derived from mouse retina tissue libraries at

various stages of development. For a detailed description

of the library from which this clone was derived, please

visit our web site at [brainest.eng.uiowa.edu](http://brainest.eng.uiowa.edu).

TAG\_SEQ=None found"

BASE COUNT 108 a 140 c 119 g 125 t

## ORIGIN

Query Match 8.1%; Score 168.6; DB 11; Length 492;  
 Best Local Similarity 61.1%; Pred. No. 2.5e-21;  
 Matches 273; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 1316 agagtgggcagccctcgccatttattctgtgcccagatgtatgcacagggag 1375  
 Db 479 AGGGAGGCCATATCTTGGCTTCTTATAACCTGCACAGATGCAGCGCACAGGG 420  
 QY 1376 tagtaagatcatgcagaactcgtgtgagctttataaaagtctctgtaactaggccact 1435  
 Db 419 TGATGGCGTCTCTACACTGCAGTGGAGTTGTTTAAAGATGTGTGAGCGGAGTCGCT 360  
 QY 1436 gggctgagaaatccctgcagacttacttgcctataagatgtgatataagattctctc 1495  
 Db 359 GGTGAGAGAGACTGATGACTGCTTACACAGCTATTAAAGATGAGACTCAATTCGTCGAG 300  
 QY 1496 ttgttcagtatgcactgcttcagaaatggggtatgaagttagctcaaaagcaattcagcat 1555  
 Db 299 TGGTCCAGTACCTCTGCTGGCTGAGCAGGGCTACGAGGTGGCGCAGAGCAACGCGAGCCT 240  
 QY 1556 tcatttggaaataaaaaaggctaacattcttgaaaaagagaagatgtatcccaatggcgc 1615  
 Db 239 TCATCCTCCAGCAGAGAGAGAACCAACCATTTGTAGTGAGAATGAAGCTTATCCCGAGAGCTT 180  
 QY 1616 ttctctatggaatcgagctgccattcaaggcaatgcatttgcctagatgaaatggag 1675  
 Db 179 TACTGCATTGGACACAGGGCGCTCCCAAGGTTTACTGTGGCTAGAAATAAGCTTTGGAG 120  
 QY 1676 attaccattactatgctatgggactaagaagactatcaaacagagccacacactaca 1735  
 Db 119 ACTACCACCTCTATGGCTTTGGCACTGATGTGATTATGAGACCGCATTTATTTCATTACC 60  
 QY 1736 gattgcagccaaataataaccacaagc 1762  
 Db 59 GCCTGGCTTCTGAGCAGCAGCAGCAGCG 33

## RESULT 12

BF727590

LOCUS

DEFINITION

BF727590 659 bp mRNA EST 08-JAN-2001  
 SWOV3MCAM51C07SK Onchocerca volvulus molting L3 larva cDNA  
 (SL96MLW-Ovml3) Onchocerca volvulus cDNA clone SWOV3MCAM51C07 5',  
 mRNA sequence.

ACCESSION

BF727590.1 GI:12045451

VERSION

EST.

KEYWORDS

ONCHOCERCA VOLVULUS.

SOURCE

ORGANISM

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 Onchocercidae; Onchocerca.

REFERENCE

AUTHORS

Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.

TITLE

Genes expressed in molting L3 larvae of Onchocerca volvulus

JOURNAL

Unpublished (1997)

COMMENT

Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853786

Email: [genomesmith.edu](mailto:genomesmith.edu)

Seq primer: pbluescript SK.

Location/Qualifiers

1..659

/organism="Onchocerca volvulus"

/strain="Kumba, Cameroons"

/db\_xref="taxon:6282"

/clone="SWOV3MCAM51C07"

/clone\_lib="Onchocerca volvulus molting L3 larva cDNA

(SL96MLW-Ovml3)"

/dev\_stage="molting L3"

/lab\_host="XLI-Blue MRF"

/note="Vector: Lambda Uni-ZAP XR; Site\_1: Eco RI; Site\_2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in



Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.  
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.  
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.  
and Ducar,M.

TITLE  
Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression

JOURNAL  
Nat. Biotechnol. 19 (5), 440 (2001) In press

COMMENT  
Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scaine@aterssys.com  
High quality sequence stop: 545.  
Location/Qualifiers  
source  
l..763  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="HT1080"  
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT      213 a    156 c    147 g    247 t

ORIGIN

Query Match                  8.0%; Score 165.6; DB 11; Length 763;  
Best Local Similarity         97.7%; Pred. No. 8.2e-21;  
Matches 168; Conservative    0; Mismatches 4; Indels 0; Gaps

QY 379 aaagaagaaaggctacttattgtgcaaacgcagtacatgggaaccttgaaaagtcatg 438  
| | | | |  
Db 505 ACATAAGAGCCCTACCTACTTTTGGCAAAAGCAGCTGACATGGGAACCTTGAAAGCATGTG 446  
| | | | |

QY 439 gagaaaaaggcgcaccctttcgctatttggaataatttggcgtgcacaatatcacagccgt 498  
| | | | |  
Db 445 GAGAANAATGGCGTCGCCGTTGCCTATTGGAANATTTTTGGCGTGCCAATAATTAACAGCACGT 386  
| | | | |

QY 499 atccaatttatgatgcccttcctggctaagaaggatcatgttaaacccccaaaacy 550  
| | | | |  
Db 385 ATCCAATTATGAGTCCTTTGGCTAAAAGAGGATCATGTAAAGCCCCAACG 334  
| | | | |

Search completed: April 8, 2002, 21:04:09  
Job time: 2391 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2002, 20:36:08 ; Search time 53.39 Seconds  
(without alignments)  
8780.848 Million cell updates/sec

Title: US-09-714-882-1  
Perfect score: 2070  
Sequence: 1 atgaagccctgtctctgtt.....ttagaatcaccatggtag 2070

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	146.4	7.1	1124	2	US-08-966-316-11
2	90.2	4.4	571	3	US-08-699-103B-6
3	80.2	3.9	7218	1	US-08-232-463-14
4	64.4	3.1	19124	2	US-08-487-826B-13
5	60.2	2.9	2447	2	US-09-014-969-14
6	59	2.9	240	1	US-08-628-417-6
7	56.2	2.7	6243	2	US-03-056-075-1
8	48	2.3	291	1	US-07-922-723A-7
9	48	2.3	291	1	US-07-799-828C-7
10	48	2.3	291	1	US-08-074-275-7
11	48	2.3	291	1	US-08-480-366-7
12	48	2.3	291	2	US-07-952-277A-7
13	47.8	2.3	4818	3	US-08-817-926-27
14	46.8	2.3	5852	1	US-07-867-106-2
15	46.4	2.2	1117	4	US-09-247-373B-33
16	46.4	2.2	37948	4	US-09-251-645-11
17	46	2.2	636	4	US-08-998-416-1137
18	46	2.2	658	4	US-08-998-416-595
19	45.8	2.2	3763	1	US-07-792-865D-1
20	45.4	2.2	1051	4	US-09-245-041-10
21	44.4	2.1	6671	1	US-08-280-443-1
22	44.4	2.1	6671	1	US-08-457-459-1
23	44.4	2.1	6671	1	US-08-555-678-1
24	44.4	2.1	6671	5	PCT-US95-02275-1
25	44.2	2.1	1813	4	US-09-071-224-3
26	44.2	2.1	6243	2	US-09-056-075-1
27	44	2.1	144	1	US-08-702-344-26

28	44	2.1	19557	5	PCT-US92-06300-1	Sequence 1, Appli
29	43.4	2.1	2223	1	US-08-257-073-4	Sequence 4, Appli
30	43.4	2.1	2946	3	US-08-968-563-6	Sequence 6, Appli
31	43.4	2.1	2946	3	US-08-969-683A-6	Sequence 6, Appli
32	43.2	2.1	1066	1	US-08-157-101A-4	Sequence 4, Appli
33	42.8	2.1	731	1	US-08-451-405A-2	Sequence 2, Appli
34	42.6	2.1	6755	3	US-08-931-999-4	Sequence 4, Appli
35	42.6	2.1	5093	1	US-08-468-036-23	Sequence 23, Appli
36	42.6	2.1	5093	1	US-08-376-843-23	Sequence 24, Appli
37	42.4	2.0	1493	1	US-08-340-820-24	Sequence 24, Appli
38	42.4	2.0	1493	1	US-08-593-535-24	Sequence 24, Appli
39	42.4	2.0	2082	2	US-08-785-310A-2	Sequence 2, Appli
40	42.4	2.0	19124	2	US-08-487-826B-13	Sequence 13, Appli
41	42.2	2.0	376	2	US-08-623-906A-18	Sequence 18, Appli
42	42.2	2.0	454	2	US-08-623-906A-6	Sequence 6, Appli
43	42.2	2.0	837	4	US-08-998-416-288	Sequence 288, App
44	42.2	2.0	1434	2	US-08-903-801-2	Sequence 2, Appli
45	42.2	2.0	1434	4	US-09-295-055-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-966-316-11  
; Sequence 11, Application US/08966316  
; Patent No. 5932445  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Reddy, Roopa  
; APPLICANT: Murry, Lynn E.  
; APPLICANT: Mathur, Preete  
; TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08966,316  
; APPLICATION NUMBER: US/08966,316  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0424 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: LEUKNOT03  
; CLONE: 1880692  
US-08-966-316-11





TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pt2gpt-Fls  
US-08-232-463-14

Query Match 3.9%; Score 80.2; DB 1; Length 7218;  
Best Local Similarity 5.7%; Pred. No. 2.5e-11;  
Matches 22; Conservative 230; Mismatches 133; Indels 0; Gaps 0;

QY 4 aagccctgtctcttaagagatattgataattcttgggtcaccaataaaactatc 63  
DB 1481 AATTACCTATCTCAAGTAGTTAAAGAGATAGAGAATTTGGTACRRRRRRRRR 1422  
QY 64 aagcagaggaacataataaaagacaaagaaagaaatgtcacacacaggtatcagt 123  
DB 1421 RR 1362  
QY 124 aacgaatacaacaatattatcacacatatattggacaaagacatctagtaatc 183  
DB 1361 RR 1302  
QY 184 aataaaagagaaatctctggagaaaagaaagaaatcaacgtataaataaagaa 243  
DB 1301 RR 1242  
QY 244 attcaataaagatatcttgagagaaataagacatttacaagaacagcagagaaa 303  
DB 1241 RR 1182  
QY 304 aatttcagatgaaggagaccagctatttaagatgggcatacaagttctccagcgtc 363  
DB 1181 RR 1122  
QY 364 aaaaacaaacaaagaaagaaag 388  
DB 1121 RRRRRRRRRRRRRRRRRRRRRRR 1097

RESULT 4  
US-08-487-826B-13  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827

GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487.826B

FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
; NAME: Israelson, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.1%; Score 64.4; DB 2; Length 19124;  
Best Local Similarity 49.4%; Pred. No. 5e-07;  
Matches 167; Conservative 0; Mismatches 171; Indels 0; Gaps 0;  
QY 49 acaattaaactatcaagcagaggaacataataaaagacaaagaaagaaatgtcacc 108  
DB 15462 AAAATGAAATATAAAAAAATTTTAAATATAAAAAAATAAAAAAAGGAGAAA 15521  
QY 109 acacagggtatcagtgacgaacatacaacaattttcacacacatttggacaaagaaca 168  
DB 15522 AATTTTAAAAAATAAATAAAAAATTTTAAATAAATAAATAAATAAATAAATAAATG 15581  
QY 169 tctagtaagttaatacaataaaagagaaatctcctggagaaagaaagaaatcaacgtaaa 228  
DB 15582 AAAAGATTATCAAAAAAATAAATAAAAAATTTTATATAAAAAAATAAATAAATAA 15641  
QY 229 ataagaataaagggaattcaaaataaagatatcttgaagagagaaataagaatcattaca 288  
DB 15642 AAAATATAAAACAAAGAGAGAAAAAATTTAAAAAATAAATAAATAAATAAATAAATA 15701  
QY 289 aagcagcagagaaatattttacagatgaagagagccagctatttaagatggcgcacag 348  
DB 15702 AAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 15761  
QY 349 gtctccagcagctctaaagcccaaaacaaagaaaga 386  
DB 15762 AATTAATAAATAATGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 15799

RESULT 5  
US-09-014-969-14  
; Sequence 14, Application US/09014969  
; Patent No. 5965397  
GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LeVallie, Edward R.  
; APPLICANT: Racie, Lisa A.  
; APPLICANT: Werberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Agostino, Michael J.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; TITLE OF INVENTION: ENCODING THEM  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5951
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-014-969-14

Query Match 2.9%; Score 60.2; DB 2; Length 2447;
Best Local Similarity 53.1%; Pred. No. 2.5e-06;
Matches 128; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 10 ttgtctctgttaagagatattgataattcttgggtgcacaaactaaactcaagca 69
DB 2207 TTTTGTACTTTAAATGTGACAAATAAACCTTTTGGGAGAAAAAARAAAAA 2266

QY 70 gaggaacataataagacaagaagaagaatgtcaccacacaggtatcagtgacgaa 129
DB 2267 AAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 2326

QY 130 atcaaacataattttcacacatatgtggacaaagacatctagtaattgtaatacaataaa 189
DB 2327 AAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 2386

QY 190 agagaaatctctcgtggagaaagagaatcaacgtataaataagataaaaggaattcaa 249
DB 2387 AAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 2446

QY 250 a 250
DB 2447 A 2447

RESULT 6
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULYSSES J.
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; US-08-628-417-6

Query Match 2.9%; Score 59; DB 1; Length 240;
Best Local Similarity 53.7%; Pred. No. 1.9e-06;
Matches 122; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 52 attaaactatcaagcagagagacataataaaagacaaaggaagaatgtcaccaca 111
DB 14 AATAAACTTTAGATAATATTTTACTAAAAAARAAAAAARAAAAAARAAAAA 73

QY 112 caggatcagtgacgaatcaacaatattttcacacatatgtggacaaagaaacatct 171
DB 74 AAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 133

QY 172 agtaatgtaatacaataaagagaaatactctcctgggagaaagaaagaaatacaacgtaaaaata 231
DB 134 AAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 193

QY 232 aqaataaaggaattcaaaataaagatatcttgaagagaaataaagaa 278
DB 194 AAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAAARAAAAA 240

RESULT 7
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
```









Db 1090 aaacaaaaaaamaaaaaa 1111

Search completed: April 8, 2002, 21:32:04  
Job time: 3356 sec

